Fuji Electric's Analytics and AI

ASANO, Takamasa* WATANABE, Takuya* SHIRAKI, Takashi*

ABSTRACT

In recent years, companies have been accelerating their efforts to promote DX. AI is one of the core digital technologies needed to promote DX. Fuji Electric has developed the basic technologies of Analytics and AI, which is a collective term for statistical analysis and machine learning technologies used for recognition, diagnosis, prediction, and optimization. For recognition technology, we developed image recognition AI using deep learning; for diagnosis technology, we evaluated five typical algorithms of unsupervised learning; for prediction technology, we focused on filter and wrapper methods; for optimization technology, we have developed a data inconsistency detection technology that checks multiple equipment statuses simultaneously.

1. Introduction

Digital transformation (DX) has become essential for corporate management in order to respond to drastic changes in the corporate business environment and to leverage data and digital technologies to provide competitive advantages for products and services. In addition, the recent spread of Internet of Things (IoT) technology has made it easier to collect a wide variety and large amount of data. Furthermore, with the development of artificial intelligence (AI) technology, there are growing expectations for the creation of new customer value and the resolution of social challenges through the utilization of large amounts of data. AI is a core technology among the digital technologies necessary for companies to promote DX. Up until now, Fuji Electric has solved a variety of challenges in the fields of industrial plants and social infrastructure through the development of analytics and AI.(1)

Specifically, in the field of factory automation (FA), we have developed a multivariate statistical process control (MSPC) technology for batch processes as a proprietary anomaly diagnosis technology. MSPC meets the needs of sophisticated maintenance management of manufacturing equipment and quality control of manufacturing processes and this technology has contributed to preventive maintenance of equipment and reduction of defective product rates. In the ever-changing field of energy supply and demand, we have independently improved our neural networks and just-in-time (JIT) prediction technologies to meet the demand for highly accurate predictions and are applying them to forecast future energy demand. Furthermore, we have combined these technologies with mathematical programming and meta-heuristic optimization techniques to automatically plan the operation of plant equipment in order to contribute to the reduction of fuel costs and CO_2 emissions. We have also been developing "explainable AI" to address the black box problem of AI. We have developed a proprietary structured deep learning (DL)*1 technology that visualizes the correlation between input and output by using a sophisticated neural network structure, and are aiming to expand the application of AI to fields that require safety and reliability, where it has been difficult to apply AI in the past.

2. Overview of Analytics and AI

Fuji Electric's analytics and AI is a collective term for statistical analysis and machine learning technologies that perform recognition, diagnosis, prediction, and optimization (see Fig. 1). Our analytics and AI technology makes it possible to recognize situations in the production site and diagnose the cause of the incidents. In addition, it can facilitate optimization based on the prediction of future conditions in order to create new value for customers.

2.1 Recognition technology

Our recognition technology is used to save labor of equipment maintenance (refer to "Text Recognition Technologies to Facilitate Technology Transfer and Information Sharing in Equipment Maintenance" on page 174) and automate visual inspection of products by applying a proprietary pre-processing technology

* Corporate R&D Headquarters, Fuji Electric Co., Ltd.

*1 Deep learning (DL): DL stands for deep learning. Deep learning is a method of learning with computers using multiple layers of neural networks that mimic human brain nerves. It is an AI algorithm used mainly for image recognition, language recognition and prediction.

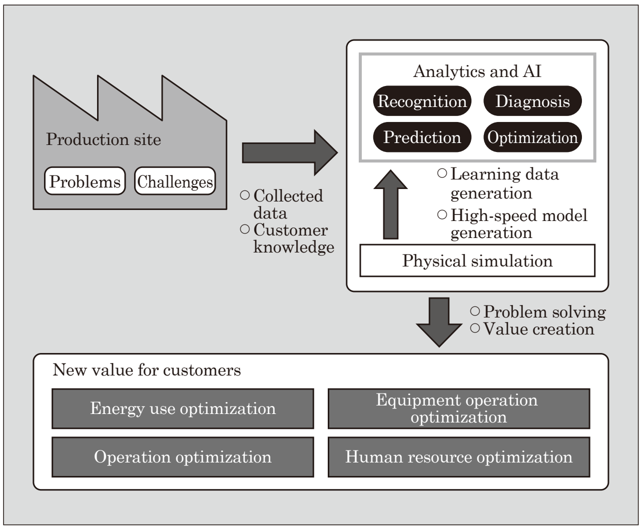

Fig.1 Overview of Fuji Electric's analytics and AI

and the latest DL technology to text and image data.

2.2 Diagnosis technology

Diagnosis technology contributes to detecting signs of anomalies in manufacturing processes and diagnosing their causes.[1] We use industry-proven MSPC for manufacturing process data whose characteristics follow a normal distribution and apply a new machine learning technique for more complex characteristics.

2.3 Prediction technology

Predicting future plant conditions can help support plant operation. We apply structured DL, capable of performing complex modeling for detailed plant operation data, to targets with large amount of data, whereas JIT prediction, capable of modeling even a small amount of data, to targets with small amount of data.[1]

2.4 Optimization technology

Optimization technology is aimed at more efficient plant operation than human operation. We apply mathematical programming when the plant has many equipment units, such as generators and boilers, or the planning period is short. We otherwise use metaheuristics.[1]

In addition, Fuji Electric is developing simulation application technologies as model building technologies to achieve digital twins by combining its simulation technology with analytics and AI

3. Introduction to Our Analytics and AI

3.1 Recognition technology

The rapid progress of AI technology in recent years has enabled the application of AI technologies, including DL technologies, to the automation of advanced human-based tasks that have been difficult to replace with conventional rule-based techniques.

In order to apply DL technologies to the industrial field, Fuji Electric has been developing elemental technologies such as a pre-processing technology to cope with insufficient training data, an anomaly detection technology that is capable of learning using only a small amount of normal data, and a visualization technology for AI decisions. Next, we will describe an application of our image recognition AI that uses DL technology to automate the visual inspection of semiconductor wafers in one of our factories.

(1) Objectives and challenges of wafer visual inspection

Fuji Electric manufactures a variety of power semiconductor products. These products are manufactured from wafers through several processes such as oxidation, pattern forming, wiring, chip forming, and mounting and packaging. Since various kinds of defects can occur in each process, visual inspections are conducted between processes, and images are taken to detect defective parts. The images are classified and counted for each mode of anomaly, and the process that caused the anomaly is identified based on the count trends of each mode. This type of approach helps improve the process (see Fig. 2). In the past, the captured images were classified visually by inspectors in the field. This classification process required a lot of time. In addition, the classification criteria differed from inspector to inspector, resulting in a high degree

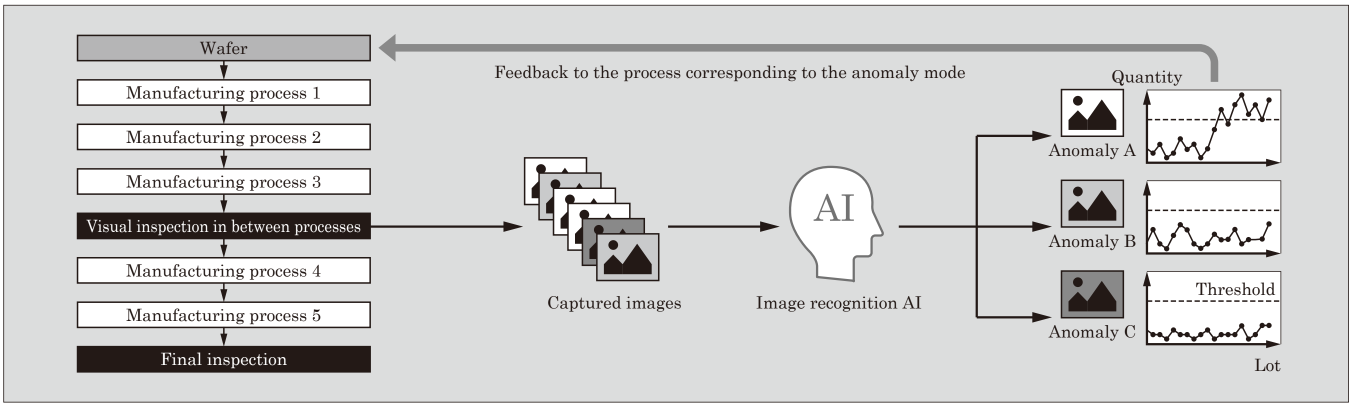

Fig.2 Application of image recognition AI to wafer visual inspection

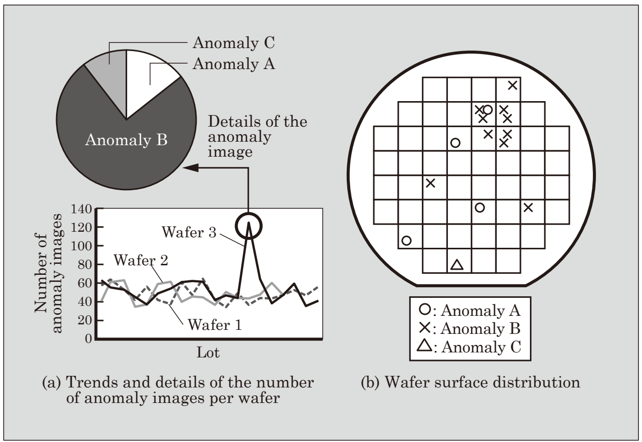

Fig.3 Example of visualizing classification results using image recognition AI

of dependency on inspector skills. It is against this backdrop that we applied image recognition AI to the classification of captured images to save labor, improve throughput, and eliminate the dependence on inspector skills in the classification and counting process.

(2) Visualization of classification results

After pre-processing the captured images as necessary by performing white balance correction, brightness normalization, and data augmentation, the image recognition AI classifies the images by anomaly mode and visualizes the results. Figure 3 shows an example of visualizing the classification results. Figure 3(a) is a graph plotting the number of images taken of each wafer in which an anomaly was found for each lot. This makes it possible to check which lot has the most wafers with anomalies. Figure 3(b) is an example of visualizing where each anomaly occurs within the wafer surface. This makes it possible to check whether the anomaly occurs uniformly within the surface or whether it is concentrated in a specific location. This approach makes it possible to analyze the tendency of anomaly occurrences from various perspectives and to identify manufacturing processes that need to be improved.

In the future, we plan to use this system to develop detection and adaptation technologies for concept drift*2, which has been an ongoing challenge of analytics and AI.

3.2 Diagnostic technology

AI-based diagnosis of anomalies in manufacturing processes has contributed to quality improvement and yield improvement. Fuji Electric has a lot of experience with MSPC, especially in the field of chemical processes. We have also developed machine learning that can accurately diagnose objects with complex characteristics, as well as functions that explain the basis for diagnosis.[1]

To diagnose complex characteristics, it is necessary to select an appropriate method using a performance evaluation technique for various applicable targets. Since the occurrence of anomalies is not so common in actual manufacturing processes, Fuji Electric is focusing on unsupervised learning, which does not require anomaly data during training. In this respect, we are experimenting with various techniques.

In this section, we will describe the results of evaluating the following five typical unsupervised learning algorithms:

(a) One class support vector machine (OCSVM): Diagnosis using a non-linear function called a kernel
(b) Isolation forest (IF): Diagnosis based on if-then rules, called decision trees
(c) Local outlier factor (LOF): Diagnosis based on distance from normal data
(d) Isolation using nearest neighbor ensembles (iNEE): Diagnosis by combining IF and LOF
(e) Ensemble K-nearest neighbor algorithm (EnsKnn): Diagnosis from multiple types of similar data

In this study, we prepared 10 sets of real data as benchmark data. It was time series data including power, temperature, and pressure, measured mainly in manufacturing processes. We built a diagnostic model by training it on the data during a normal period and verified whether the model could be used to diagnose normal or anomaly data during the period of verification. Figure 4 shows a comparison of the diagnostic performance, called the F1 score, of different machine learning methods. The F1 score is the harmonic mean of the precision (the rate of correct judgments among those judged to be positive) and the recall (the rate of correct judgments among those actually positive), where the closer the score is to 1, the better the diagnostic performance. The algorithm with the best diagnostic performance was, for example, LOF for dataset D01 and EnsKnn for dataset D02 among the results

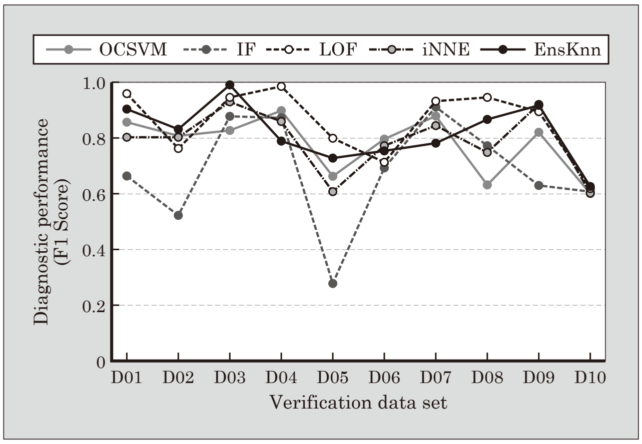

Fig.4 Comparison of diagnostic performance using machine learning techniques

*2 Concept drift: A change in the statistical properties of a target variable that an AI model is trying to predict over time due to various reasons.

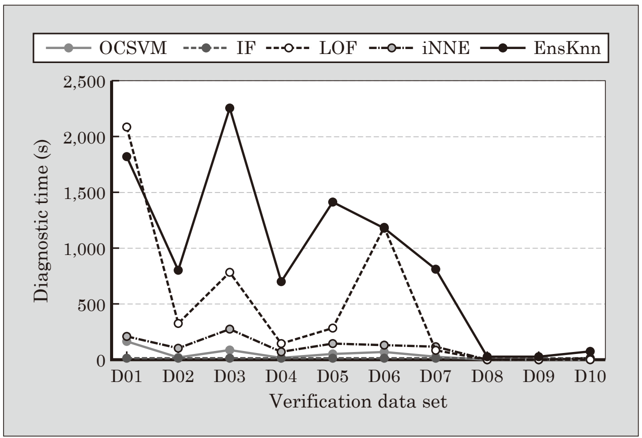

Fig.5 Comparison of diagnosis time using machine learning techniques

in Fig. 4. The best algorithm differed for each data-set. Figure 5 shows a graph comparing the diagnosis time for each machine learning technique. LOF and EnsKnn had high diagnostic performance but required a long diagnostic time.

Based on the results of this study, we confirmed that the best algorithm differed for each data set and that the diagnostic time varied depending on the algorithm. Using the F1 score and diagnostic time, it is possible to select the appropriate algorithm according to the required specifications of the diagnostic target. We are also planning to incorporate these algorithms into tools that can be easily handled by engineers and data scientists alike.

3.3 Prediction technology

Fuji Electric is developing technologies to predict energy demand and quality in order to support plant operations. We have developed some unique prediction models, such as a JIT prediction, which can make predictions using small amounts of training data, and a structured DL model, which is capable of explaining prediction results. In addition to the development of these prediction models themselves, it is important to know which input variables to choose in order to generate accurate prediction models. Normally, the selection of input variables requires several days or months of consideration, as the data scientist repeats trial and error to gain knowledge of the prediction target. We have developed some techniques to improve this. For example, for JIT prediction(1), we developed a proprietary variable selection technique that uses variable importance to reduce trial and error time.

In addition, we have focused on the filter method and the wrapper method, which are general-purpose variable selection methods that can be applied to various types of machine learning applications (see Fig. 6). The filter method can use input variables each of which the importance is above a certain value. The wrapper method looks for the combination of input variables with the highest prediction accuracy based

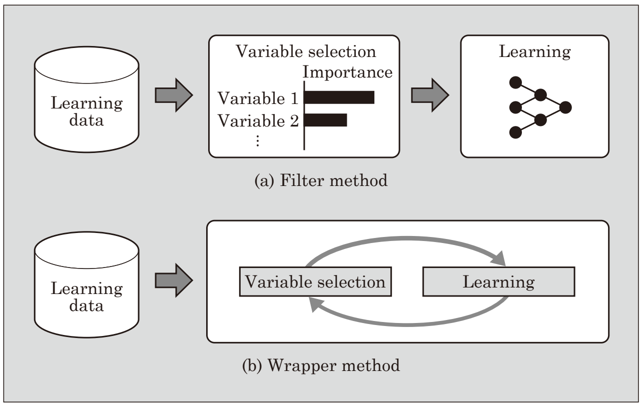

Fig.6 Conceptual diagram of the filter and wrapper methods

Table 1 Variation of prediction error according to variable selection method

	Number of input variables	Mean prediction error
No variable selection	429	4.4%
Filter method	23	3.7%
Wrapper method	61	3.1%

on the change in prediction error.

In order to verify the effectiveness of each technique, we prepared benchmark data for the task of predicting temperature several hours ahead of a manufacturing process. The candidate input variables consist of all 429 types of measured data. We used the partial least squares (PLS) regression as the prediction technique because it makes it easy to handle many input variables.

Table 1 shows the prediction results. Both the filter and wrapper methods were able to reduce the mean error using only a few input variables compared to methods without variable selection. Furthermore, the wrapper method had a smaller mean error than the filter method.

In the future, we plan to apply this variable selection technology to actual products such as plant prediction support services and also to learning tools for anomaly diagnosis.

3.4 Optimization technology

The optimization technology that has been conventionally applied to energy management systems (EMSs) creates plant models of utility equipment in plants and buildings and seeks optimal equipment operations (optimal combinations) that minimize fuel costs and CO_2 emissions. Based on the optimal solution obtained, it provides guidance to operators and automatically controls the utility equipment.

To obtain the optimal operation of the plant, the plant model is first formulated as a mixed integer programming problem. In addition to this, it is necessary to provide various input and constraint conditions to facilitate prediction function demand forecasting, operator operation planning, and equipment characteristics. After satisfying these conditions, it obtains the optimal solution using mathematical programming. In the case of large-scale power systems or power systems that contain models that cannot be formulated mathematically, it applies meta-heuristics such as particle swarm optimization (PSO) to find the optimal solution.

During optimization calculations, if there are inconsistencies in the input conditions or if they are accidentally not set, the constraints will be violated and the calculation will terminate abnormally. In the event that the calculations terminate normally, it may still be the case that the calculation results contain abnormalities, such as overestimation of the energy-saving effect. When this happens, a system engineer needs to examine all the input conditions carefully to find out where the error occurred. In addition, since errors are not limited to a single location, it is necessary for the engineer to check the results repeatedly, increasing labor-hours even more (see Fig. 7).

Therefore, we have solved this problem by automatically identifying inconsistent data and indicating where to correct the data when there is abnormal termination to the calculation or an abnormal result in the calculation (see Fig. 8). Furthermore, we have implemented the following measures to identify which constraint conditions resulted in no solution, enabling the cause of the abnormal termination to be understood more easily:

(1) Equipment-specific constraint checks

When optimizing equipment with various constraints such as upper and lower fuel input limits and upper and lower output limits, abnormal termination will occur if the settings are such that calculations cannot be performed, such as when settings for operation plans exceed the upper and lower limits or when the upper and lower limits are set to be reversed wrongly. When abnormally terminated, it switches between valid and invalid constraint conditions and performs recalculation, and when the calculation becomes possible, it determines that the invalid constraint condition was the cause of the anomaly.

(2) Equipment-to-equipment constraint check

If there are constraint violations among multiple pieces of equipment, it might not be possible for it to make a decision by checking the constraint conditions for each piece of equipment as described above. For example, assuming that there is a supply and demand relationship (supply and demand constraint) for energy where the output of equipment A and B is the input of equipment C, as shown in Fig. 9. For example, if stop instructions are given to equipment A and B, and output instructions are given to equipment C, checking equipment A, equipment B, and equipment C individually will not violate the constraints in their respective equipment plans, but the plant as a whole will violate the constraints, resulting in abnormal termination. This is a case in which something that appears to be normal at first glance becomes abnormal when equipment are combined into an energy network model.

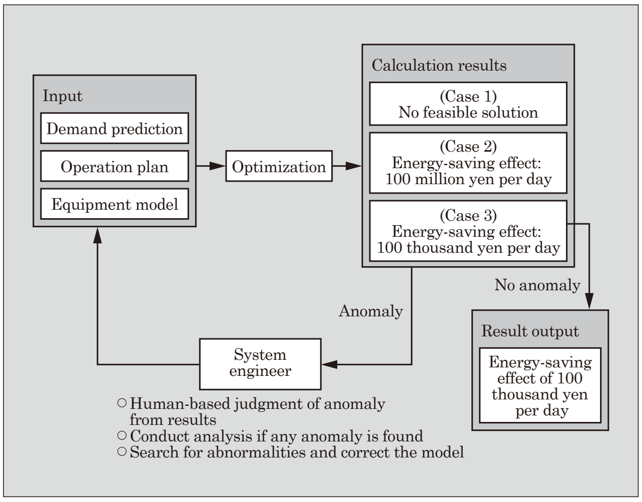

Fig.7 Conventional engineering

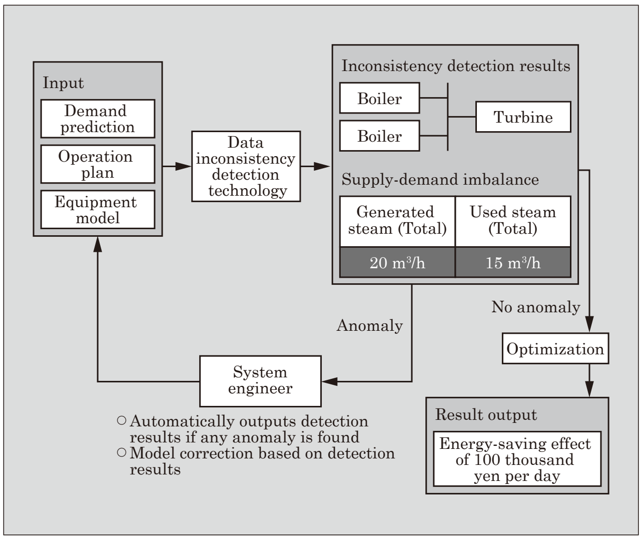

Fig.8 Engineering applying data inconsistency detection technology

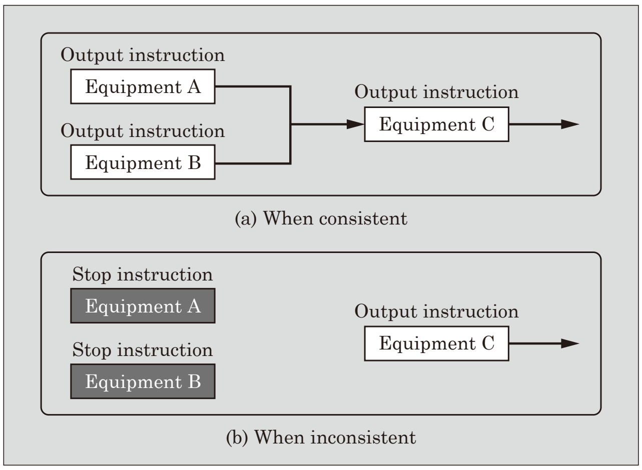

Fig.9 Examples of consistency and inconsistency in equipment-to-equipment constraints

This means that it becomes difficult to detect the anomalies by just checking the usual numerical values. To solve this problem, we have developed and are using a data inconsistency detection technology that searches for abnormal supply and demand constraints by analyzing the energy network model and switching between valid and invalid supply and demand constraints, while also checking the status of multiple pieces of equipment simultaneously.

In the past, it was necessary to be familiar with energy network models and optimization techniques in order to find the cause of abnormal termination when simulations were performed for equipment replacement or configuration changes. However, this data inconsistency detection technology will shorten engineering data analysis processes and support user maintenance tasks, without requiring any specialized knowledge.

3.5 Simulation technology for realizing digital twins

A digital twin is a technology that reproduces the functions and operations of real equipment and products in digital space, and links them in real time with operational data in real space. By using a digital twin, the present and future operating conditions of equipment and products can be grasped in real time, enabling the creation of customer value such as reduced maintenance costs through preventive maintenance of equipment and energy saving by maintaining optimal operating conditions.

Fuji Electric has been developing simulation technology focusing on structural design. By integrating analytics and AI into the simulation of physical behavior under various control and environmental conditions performed on digital devices, we can expect to improve the efficiency of the product life cycle, including not only product design but also testing and maintenance, and even apply it to the realization of digital twins.

In this section, we will discuss simulation technologies for the realization of digital twins, such as parameter identification technology to improve the accuracy of simulations, and surrogate model conversion technology to enable conventional simulations to be performed more quickly and in real time.

(1) Parameter identification technology

Depending on the design information, the simulation includes groups of parameters whose values are known (e.g. dimensions) and groups of parameters whose values are unknown (e.g. degradation state). In particular, the groups of parameters whose values are unknown are one of the factors that reduce the accuracy of the simulation of real phenomena.

Therefore, we developed a technology that applies our AI based optimization technology to identify those parameter groups for optimal values. The accuracy of the simulation was improved by identifying groups of parameters that minimize the error between the simulation output results and the actual system's collected data.

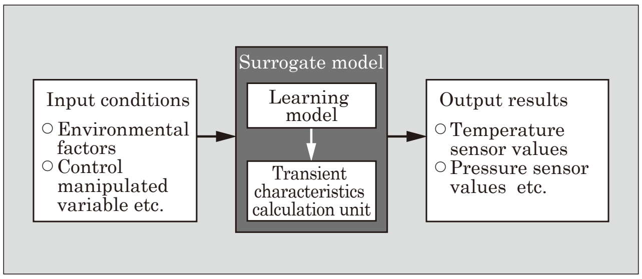

Fig.10 Surrogate model configuration

(2) Surrogate model conversion technology

Simulations can take a long time because of the huge amount of calculations that must be performed. This means that it was not possible to satisfy the stable real-time performance required for testing and maintenance applications (i.e., performance that always finishes calculation within a predetermined time, which is generally 5 seconds).

Therefore, by applying machine learning and optimization techniques, we developed a technology to convert a model into a surrogate model that completes calculations in less time than normal simulations.

Figure 10 shows the structure of the surrogate model described above. The surrogate model consists of a learning model for the steady-state characteristics and a calculation component for the transient characteristics. The learning model is generated using machine learning such as DL after creating training data by running the simulation input parameters (such as environmental factors and control manipulated variable) under multiple conditions to obtain output results (such as temperature sensor values). The transient characteristics calculation component consists of an equation with parameters such as time constants. The optimization technique determines the optimal values of those parameters so that the error between the model output and actual data is minimized.

The surrogate model enabled the simulation to be sped up (computation time: within 1 second), instead of the previously slow and unstable time (computation time: 10 seconds to 5 minutes).

Next, as an application example of this technology, we will show a HILS*3 configuration, as a simulation that aims to efficiently verify showcase controller operation without actually installing a showcase.

Figure 11 shows the configuration of the HILS for the showcase. First, we generated a surrogate model that predicts the sensor values at various locations, such as the temperature inside the showcase, and installed it on a PC as a virtual device. We were able to verify the operation of the control microcomputer by connecting the PC to it. This not only eliminated the

*3 HILS: HILS, which stands for hardware in the loop simulation, is a simulator for verification that reproduces the behavior of controlled objectives on PC.

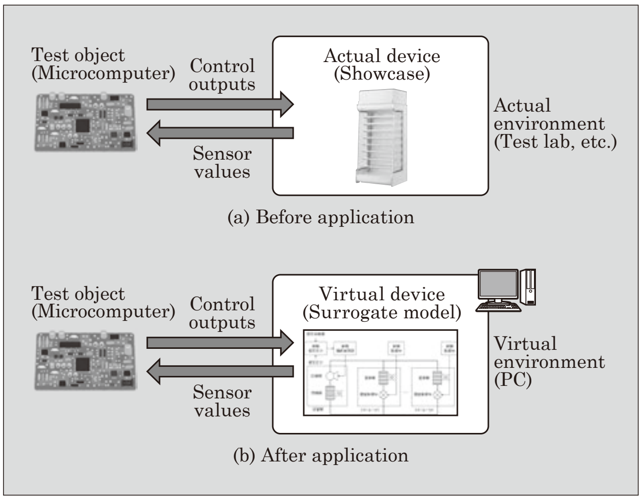

Fig.11 Configuration of HILS for showcases

need for testing using actual equipment and environments, which used to require large costs and labor-hours to build an environmental test lab and to switch test conditions, but also made it possible to reproduce all seasonal conditions, such as those of summer and winter, as well as various operating and failure conditions. This innovation will enable faster development of showcases with even greater energy-saving performance and showcases that use environmentally friendly refrigerants.

4. Postscript

In this paper, we discussed the analytics and AI that are at the core of Fuji Electric's DX.

In order to expand the application of analytics and AI, it is necessary to enhance the elemental technologies in each process of AI development, such as conceptualization, proof of concept (PoC), implementation, and operation.

Moving forward, Fuji Electric will continue to contribute to the creation of new customer value and the resolution of social challenges by accelerating the development of elemental technologies for analytics and AI and realizing digital twins.

References

(1) Iizaka, T. et al. Data Analytics as Core of Value Creation. FUJI ELECTRIC REVIEW. 2018, vol.64, no.3, p.148-153.

Fuji Electric

Innovating Energy Technology