Data Modeling Technology in Railway Operation and Maintenance

The rail industry has been analyzing large volumes of data generated daily in areas such as passengers, facilities, and rail services in recent years. Its use is being studied for improving user services and transport quality. Data can be linked by associations, correlations, and causation. Inferring causal relationships among data enables more effective anomaly detection and cause inference. This article presents an approach used to develop a cause and effect modeling method, along with technology used to infer causal relationships.

Kojin Yano, Ph.D.
Tetsushi Suzuki
Kenichiro Okada
Wei Wang, Ph.D.
Taisuke Takayanagi

1. Introduction

The rail industry uses data for transport, administration, maintenance, and various other activities, and demand for the use of data in maintenance activities has recently been growing. Rail industry maintenance is done mainly on facilities and rolling stock, with inspection and replacement basically carried out at predetermined intervals specified by time or travel distance. Procedures designed to ensure safety and stable transport have previously ensured safety by analyzing past failures to determine inspection and replacement cycles and items. But technological advances have improved the reliability of devices used in facilities and rolling stock, creating the potential for excessively frequent inspection and replacement cycles and calling for the rightsizing of maintenance work while still ensuring safety.

The rail industry is facing problems such as a growing risk of accidents from aging facilities and labor cutbacks in operations caused by the mass retirement of experienced employees. These problems call for existing maintenance systems to be improved by adapting them to the current conditions instead of continuing to use them as-is. The recent rise in the Internet of Things (IoT) is enabling various types of data to be acquired from devices in ground facilities and rolling stock, so there are calls for the use of new data-driven operation methods as a way to solve problems.

2. Rail Industry Work on Data Use

Most of the data acquired by the rail industry for facility and rolling stock maintenance is recorded chronologically in the form of facility/device operation logs. Rolling stock device operation data includes items such as time information, distance traveled, device operation count, operation time, control command information, and control results. The data obtained from rolling stock devices is used for various

Figure 1 — Data Use Applications

Data use applications are classed as either visualization or handling applications. Data first needs to be visualized to reveal how to handle it. Accumulating handling results reveals appropriate handling methods.
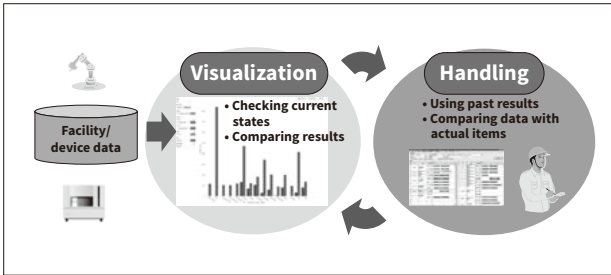
applications that are generally classed as either visualization applications or handling applications.

Visualization applications can detect anomalies from data by methods such as comparing control command information/counts and command results by train or by distance traveled, and generating alarms when abnormal operations are found. Handling applications use data when setting handling policies and in similar circumstances. For example, a handling policy might call for replacement of a device if it exhibits anomalous operation five times or more within a fixed time interval (see **Figure 1**).

Increasing rail maintenance efficiency requires facility and rolling stock failure precursor detection. But since the rail industry has always focused on ensuring transport safety and stability and there are very few past examples of failures, monitoring just the states of single devices is very unlikely to produce any major gains in improvements. To make progress in the use of data, there is therefore a need for early detection of unusual conditions by identifying and modeling the relationships among data.

3. Analysis Technology Overview and Proposed Approach

The type of data analysis under discussion makes heavy use of statistical techniques and machine learning. Statistical techniques include regression analysis, clustering, and factor analysis. Machine learning uses models such as support vector machines (SVMs) and deep neural networks (DNNs).

But while the models inferred using these methods accurately describe past trends, they are often difficult for humans to interpret. During actual data analysis, model inference is also often made difficult by inadequate data gathering or by noise, omissions, or other anomalies in the data itself.

Hitachi has therefore created an approach that progressively models relationships among data both statistically and causally, instead of just statistically. The approach is designed to produce models that enable easy interpretation by humans. **Figure 2** shows an example of a conventional modeling method and this cause and effect modeling method.
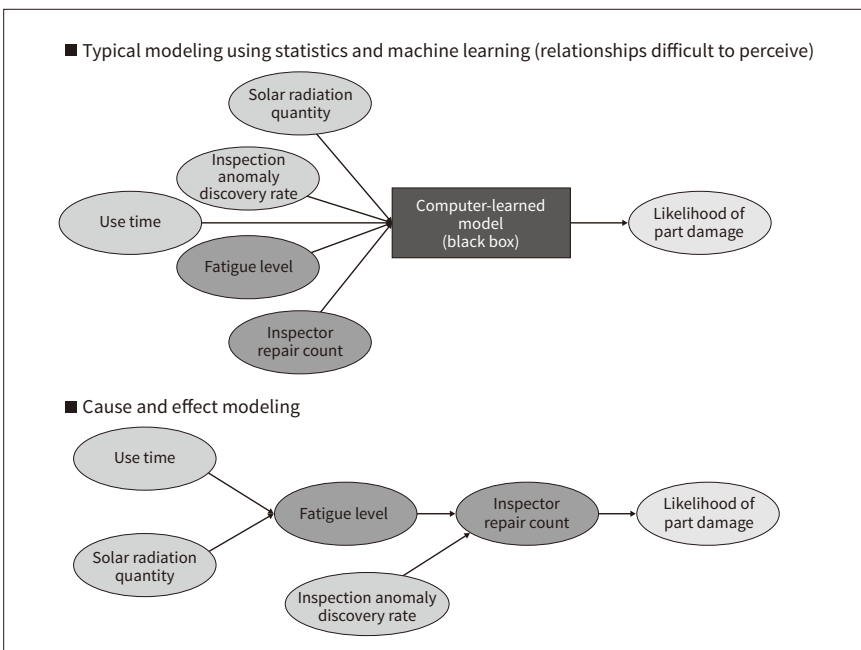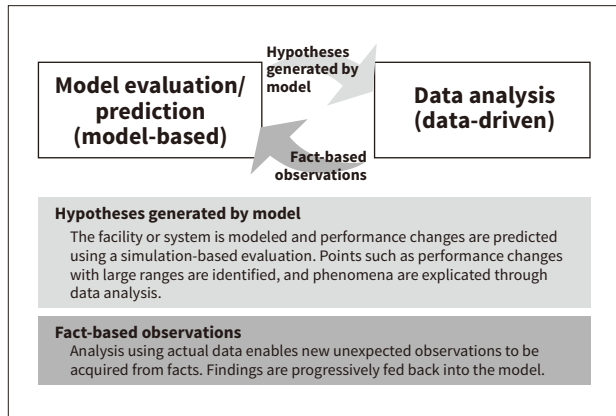

Figure 2 — Comparison of Conventional Modeling to Cause and Effect Modeling

The diagram compares the conventional modeling method to Hitachi's cause and effect modeling method. The conventional method makes it difficult to perceive data relationships.

Once the model inputs and outputs have been set, the conventional method infers a model that expresses their relationship, but the inferred model is often a black box that makes it difficult to understand the relationship. In contrast, this cause and effect modeling method progressively expresses causes and effects, producing a model that is easy to understand by humans. For example, the causal relationships output in **Figure 2** show that the likelihood of parts damage decreases when the inspector repair count becomes high and increases when it becomes low. In turn, the inspector repair count is determined by the inspection anomaly discovery rate, and fatigue level during inspection. The fatigue level is shown to be related to parts' usage time and solar radiation quantity. Progressively expressing causal relationships in this way makes it easy to infer which improvements will be most effective.

To analyze data, Hitachi has proposed an approach that combines both data-driven and model-based techniques instead of just using data-driven techniques alone. With the former type of techniques, the model is inferred from the data alone. With the latter type, accumulated design knowledge and operations knowledge are used as the basic model and incorporated into the analysis. This approach should produce accurate models faster than the conventional approach (see **Figure 3**).

Effectively implementing this approach calls for models obtained by data analysis that are easy to interpret by humans. The cause and effect modeling method enables effective application of this approach. Specifically, data analysis is used to extract causal relationships among data. An analyst then verifies these relationships to enable an assessment of the model's validity, and to enable expertise to be modeled and incorporated into the causal relationships. The constructed model is then subjected to repeated verification with data to make it more detailed and accurate.

4. Causation Inference Technology Assisting Data Modeling

4. 1
### Causation Inference Technology Overview

To enable efficient cause and effect modeling using the method described above, Hitachi has developed technology for inferring causation from data. **Figure 4** shows an overview of the technology.

The input information is composed of data from sensors that measure devices, and inspection histories at the time of measurement. One variable of interest to be used as the maintenance repair and replacement standard is specified from among this input information, and the hierarchical structure of its causal relationships is automatically inferred. This process enables the data analysis shown in **Figure 2** to rapidly identify the data relationship of interest and promote fact-based observations.

Causal relationships can take physical, stochastic, or various other forms. The technology developed is intended mainly for analysis of civil engineering or

Figure 4 — Causal Relationship Inference Algorithm Overview
The hierarchical structure of causal relationships is automatically inferred when one variable of interest is specified from among the input information.
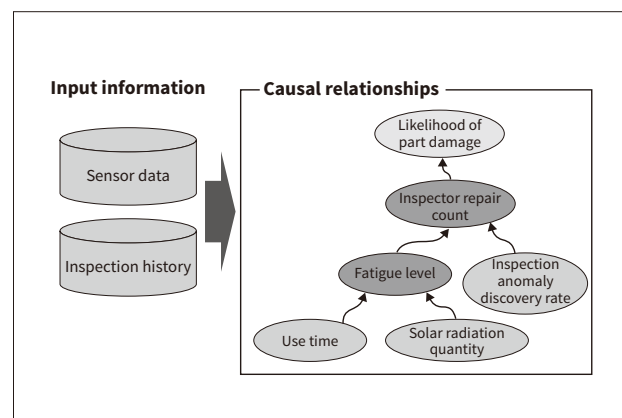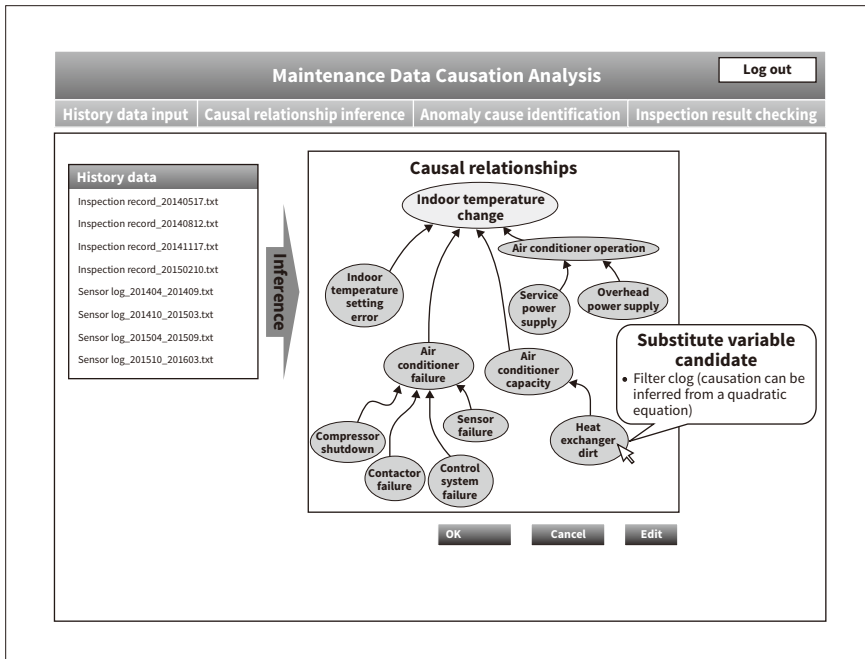
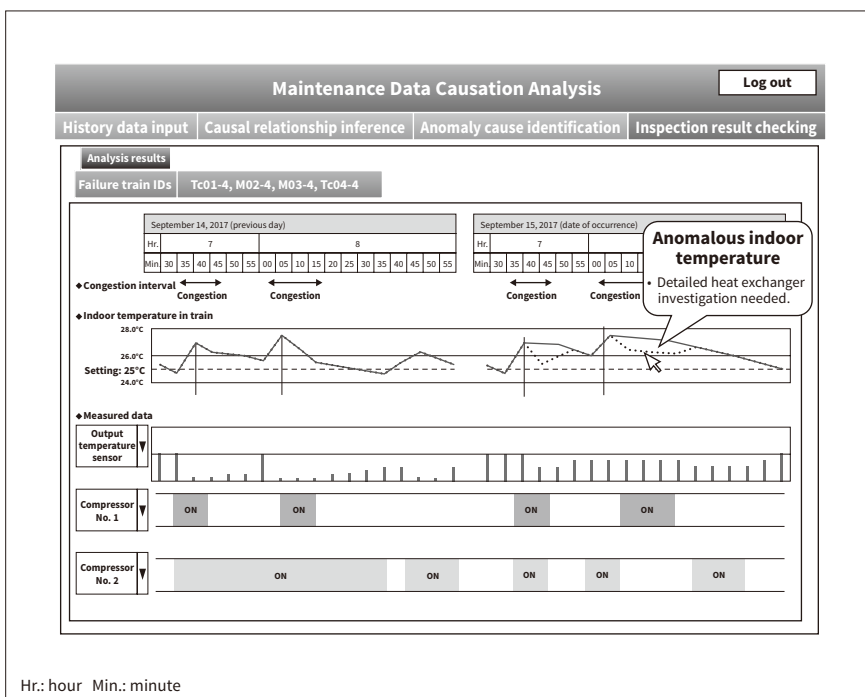**Figure 5 — Maintenance Application Example of Causal Relationship Learning**
The diagram illustrates how learning from history data takes place. For maintenance applications, causal relationships between device anomalies are inferred from device sensor data and inspection data.

mechanical equipment, and focuses on finding physical relationships. It therefore expresses causal relationships as polynomials, has explanatory variables composed of polynomials that are used either as-is or through differentiation/integration, and extracts causal relationship structures with high levels of significance from among numerous combinations of these explanatory variables.

4. 2

Application to Rail Industry Maintenance

This section describes rail industry air conditioner maintenance as an example to illustrate how identified causal relationships are used in operations. **Figure 5** illustrates the process used to infer causation from data. The left side of the diagram is the data selection area. The right side shows the results that the cause and effect model has automatically inferred from the

Figure 6 — Maintenance Application Example of Anomaly Detection Using Causal Relationships
The diagram illustrates the use of a cause and effect model to find an anomaly. Deviation is found between the actual indoor temperatures and the indoor temperatures inferred from a model produced from causal relationships, and the origin point of the deviation is inferred.

selected data. Sometimes even maintenance staff may not be able to explain the structure of the output causal relationships since the data may contain noise/omissions, or may contain data exhibiting the same movements (collinearity). In this case, it should be possible to construct explainable causal relationships by adding intermediate automatic causation inference processes, and expressing second or third candidates that were previously not used as causations. As shown in the diagram, moving the cursor to a variable that was selected once displays a second candidate variable name and causal relationship. Analysts can revise models to make them more accurate by using their expertise to change variables and causation formulas.

The following is a description of how inspection operations are done using constructed cause and effect models. **Figure 6** illustrates how a device anomaly is discovered using a cause and effect model.

The graph in the center of **Figure 6** represents air conditioner temperature. The solid line shows the actually observed air conditioner temperature, and the dotted line shows the air conditioner temperature predicted by the cause and effect model. Indoor temperature in trains is affected by the passenger boarding rate and outdoor air temperature, making it difficult to detect air conditioner anomalies from temperature changes alone. But using a cause and effect model makes it possible to estimate temperature changes

from the operation of the compressor, heat exchanger, power supply, and other devices in the air conditioner. Comparing the estimated values to the actual indoor temperatures enables identification of air conditioner anomalies that would not be immediately apparent from temperature changes alone.

Finally, the following describes the operations done to infer the locations of anomalies. **Figure 7** illustrates how the location of an anomaly is inferred using a cause and effect model. The right side of the diagram shows the causal relationships. The state of the model shown here indicates that previous analysis has found temperature changes that were assessed as anomalous. Tracing backwards through the causal relationships should make it easy to infer the location of the anomaly. Specifically, multiple explanatory variables that have a causal relationship with the variable of interest (temperature change) are searched to identify the explanatory variable that has actual values different from the values inferred by the cause and effect model [**Figure 7 (1)**]. The explanatory variables that have a causal relationship with explanatory variable (1) are then searched to identify the explanatory variable with anomalous values [**Figure 7 (2)**]. By assisting with inferring the location of anomalies in this way, cause and effect modeling should reduce maintenance operation workloads and ensure continued maintenance quality.
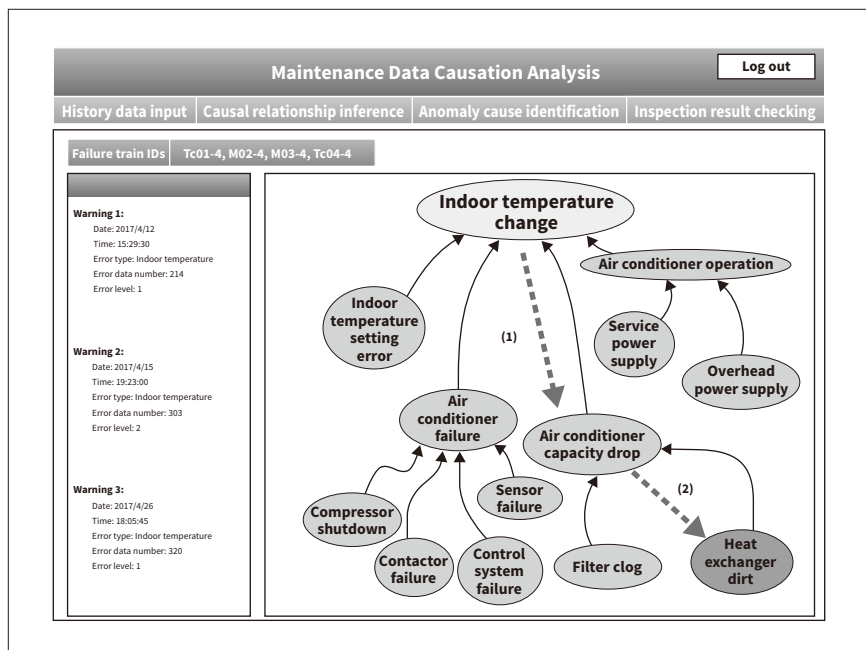
Figure 7 — Maintenance Application Example of Cause Analysis From Causal Relationships

The diagram illustrates the use of a cause and effect model to infer the location of an anomaly. Causes are identified by tracing the variables with differences between actually measured values and values inferred by the model.

5. Conclusions

This article has described how the rail industry is using data for maintenance applications. Modeling that looks at data relationships will be key for making more effective use of data and progressing beyond conventional assessments based on the states of single devices. Hitachi is therefore working on developing cause and effect modeling technology, and has shown here how this technology can be applied. Hitachi plans to provide solutions to the rail industry while applying its technology to actual data to accumulate a successful verification track record.

References

1) S. Kubo, "Innovations in Railway Maintenance Technology Using ICT," 29th Railway Technical Research Institute Lecture (Nov. 2016), in Japanese.

2) H Morimoto et al., "Hitachi Rail Innovation: Future Rail Services Driven by Digital Technologies," Hitachi Review, 67, pp. 440-447 (Jun. 2018).

3) H. Nagaoka et al., "Development of Methods for Visualizing Customer Value in Terms of People and Management," Hitachi Review, 65, pp. 840-846 (Mar. 2016).

Authors

Kojin Yano, Ph.D.
Social Systems Engineering Research Department, Center for Technology Innovation – System Engineering, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research on transportation system modeling and simulation technology. *Society memberships*: The Institute of Electrical Engineers of Japan (IEEJ), the Society of Socio-Informatics (SSI), the Information Processing Society of Japan (IPSJ), and the Operations Research Society of Japan (ORSJ).

Tetsushi Suzuki
Transportation Systems Department 4, Transportation Information Systems Division, Social Infrastructure Information Systems Division, Social Infrastrucre Systems Business Unit, Hitachi Ltd. *Current work and research:* Development of railway maintenance and asset management systems.

Kenichiro Okada
Transportation Systems Department 4, Transportation Information Systems Division, Social Infrastructure Information Systems Division, Social Infrastrucre Systems Business Unit, Hitachi Ltd. *Current work and research:* Development of railway information and control systems. *Society memberships*: The Society of Project Management (SPM).

Wei Wang, Ph.D.
Social Systems Engineering Research Department, Center for Technology Innovation – System Engineering, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research of new data analytics technology for smart railway maintenance.

Taisuke Takayanagi
Electromagnetic Application Systems Research Department, Center for Technology Innovation – Energy, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research on modeling and simulation technology for social infrastructure.