

Title:

DATA COMPRESSION DEVICE, DATA COMPRESSION METHOD, AND DATA
COMPRESSION PROGRAM

5 Technical Field

[0001] The disclosure relates to a data compression device, a data compression method,
and a data compression program.

Related Art

[0002] There are cases where a machine learning model such as a neural network
10 including a fuzzy neural network learns a relationship in time series data collected by a
sensor or the like installed in a chemical plant or the like, and the model that has performed
learning is used to perform anomaly sign detection or continuous calculation of
characteristic values that are difficult to measure.

[0003] For example, in the case of anomaly sign detection, learning for constructing a
15 model for detecting anomalies is performed. As a conventional technology, a technology
is known in which an evaluation value of an anomaly detection model is calculated and the
anomaly detection model is efficiently created (for example, see Patent Document 1).

Prior Art Documents

Patent Documents

20 **[0004]** Patent Document 1: Japanese Patent Laid-Open No. 2022-056746

SUMMARY

Technical Problem

[0005] However, in the above-mentioned conventional technology, there are cases
25 where it is difficult to accurately perform model learning. For example, the conventional

technology is a method in which a nonlinear regression model is constructed and is used in anomaly detection. However, in the conventional technology, since data obtained from a sensor or the like is used as it is during construction of the nonlinear regression model, if time series data having a biased data distribution is used as learning data, it is difficult to accurately perform learning.

Solution to Problem

[0006] Accordingly, to solve the above-mentioned problem and achieve an objective, a data compression device of the disclosure is characterized by including: an acquisition part, acquiring learning data serving as a target from predetermined time series data; a division part, dividing the learning data acquired by the acquisition part based on an explanatory variable included in the learning data, and constructing a subspace; and a compression part, compressing the learning data for each of the subspace constructed by the division part, and generating compressed learning data used for learning a regression model.

Effects of Invention

[0007] According to the disclosure, an effect is achieved that accurate model learning is facilitated.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIG. 1 is a diagram describing an overview of processing by a data compression device according to the present embodiment.

FIG. 2 is a diagram describing time series data having bias.

FIG. 3 is a diagram showing an example of a configuration of the data compression device according to the present embodiment.

FIG. 4 is a table diagram showing an example of time series data according to the present embodiment.

FIG. 5 is a table diagram showing an example of learning data according to the present embodiment.

FIG. 6 is a table diagram showing an example of compressed learning data according to the present embodiment.

5 FIG. 7 is a diagram showing an example of data compression processing according to the present embodiment.

FIG. 8 is a diagram showing an example of data compression processing according to the present embodiment.

FIG. 9 is a flowchart showing an example of a procedure for data compression processing
10 according to the present embodiment.

FIG. 10 is a hardware configuration diagram showing an example of a computer that realizes a function of the data compression device according to the present embodiment.

DESCRIPTION OF THE EMBODIMENTS

15 **[0009]** Hereinafter, implementation modes (hereinafter referred to as “embodiments”) will be described with reference to the drawings. In the following description, common components are given the same reference numerals, and repeated descriptions are omitted. The description of the embodiments given below does not limit the data compression device, data compression method, and data compression program according to the disclosure.

20 **[0010]** <Introduction>

First, an introduction to the present embodiment is given. FIG. 1 is a diagram describing an overview of processing by a data compression device 100 according to the present embodiment. The data compression device 100 shown in FIG. 1 is an example of a computer that provides a technology for, after constructing a subspace with respect to

learning data (hereinafter sometimes simply referred to as “learning data”) used for model learning, compressing data for each of the subspace.

[0011] (Background)

A technology is known in which an anomaly sign is detected or characteristic values that are difficult to measure are continuously determined using a regression model that has learned a relationship in time series data collected by a sensor installed in a chemical plant or the like.

[0012] In the case of detecting an anomaly sign or the like, a predetermined regression model used in such detection is learned in advance. Accordingly, a reference technology is known in which, by calculating and evaluating an evaluation value of a nonlinear regression model, the nonlinear regression model to be learned is efficiently constructed.

[0013] However, in the reference technology, since the time series data collected by the sensor or the like is used as it is during construction of the nonlinear regression model, if there is a biased data distribution or the like, it is difficult to accurately perform learning.

[0014] Here, an example of time series data having bias is described. FIG. 2 is a diagram describing time series data having bias. FIG. 2 shows time series data ((1) of FIG. 2) collected by a sensor or the like installed in a chemical plant or the like, and a nonlinear regression space ((2) of FIG. 2) constructed using the time series data.

[0015] Chemical plants or the like tend to have a long period of a steady state in which operation is performed at a constant load and a short period of a non-steady state in which an operation such as switching loads is performed. Hence, there are cases where time series data as shown in (1-1) of FIG. 2 is collected by a sensor installed in a chemical plant. In the time series data shown in (1-1) of FIG. 2, approximately 90% of the entire time series data is collected in the steady state, and the remaining approximately 10% is collected in the non-steady state.

[0016] Here, in the case of constructing a regression model that predicts an output sensor value (objective variable) shown in (1-2) of FIG. 2 from an input sensor value (explanatory variable) shown in (1-1) of FIG. 2 in the chemical plant, the space of the constructed regression model becomes a graph with the horizontal axis as input and the vertical axis as output, as shown in (2) of FIG. 2.

[0017] As described above, in the time series data collected in the chemical plant, between the steady state and the non-steady state, there are cases where data exists with bias toward a portion of the steady state. Hence, in the space of the regression model shown in (2) of FIG. 2, data pertaining to the steady state exists with bias ((2-1) and (2-2) of FIG. 2).

[0018] In a method for constructing a general regression model such as a neural network, since each point of learning data is evaluated as being of the same importance, if the regression model is constructed using data having a biased distribution as it is, the regression model may be affected by the bias. For example, in the above-mentioned time series data collected in the chemical plant, there are a large amount of data pertaining to the steady state and a small amount of data pertaining to the non-steady state. Thus, learning is performed with excessive emphasis on the portion of the steady state, leading to problems such as reduced accuracy in a portion of the non-steady state.

[0019] As the learning data increases, computational resources pertaining to learning, such as calculation cost and memory usage, become necessary, and efficient learning becomes difficult.

[0020] (Overview of Processing by Data Compression Device 100)

Accordingly, in the data compression device 100 according to the present embodiment, with respect to the acquired learning data, the space of the regression model composed of the objective variable and the explanatory variable is constructed into a subspace along an axis of the explanatory variable, and the learning data included in the subspace is compressed.

[0021] Here, referring back to FIG. 1, a sequence of processing by the data compression device 100 is described. First, in the data compression device 100, learning data serving as a target is acquired from predetermined time series data that is collected by an external sensor or the like and stored in a storage part 120 ((1) of FIG. 1).

5 **[0022]** In the data compression device 100, based on an explanatory variable included in the acquired learning data, the learning data is divided and a subspace is constructed ((2) of FIG. 1). Next, in the data compression device 100, the learning data for each constructed subspace is compressed, and compressed learning data (hereinafter sometimes simply referred to as “compressed learning data”) used for learning the regression model is
10 generated ((3) of FIG. 1).

[0023] As described above, in the data compression device 100 according to the present embodiment, by utilizing the fact that the learning data has bias in the space of the regression model, data volume can be compressed while information for learning the regression model is held. Accordingly, in the data compression device 100, the compressed learning data
15 enabling efficient (with low computational resource usage) and highly accurate construction of the regression model can be generated. Accordingly, in the data compression device 100, an effect is achieved that accurate model learning is facilitated.

[0024] <Description of Data Compression Device 100>

(Data Compression Device 100)

20 Hereinafter, detailed functions of the data compression device 100 according to the present embodiment will be described. FIG. 3 is a diagram showing an example of a configuration of the data compression device 100 according to the present embodiment. As shown in FIG. 3, the data compression device 100 includes a communication part 110, the storage part 120, and a control part 130.

25 **[0025]** The data compression device 100 includes an input part (not shown) such as a

keyboard or touch panel for receiving input from a user or the like, and a display part (not shown) such as a display or printer for displaying to the user or the like a result of information processing by the data compression device 100.

[0026] (Communication Part 110)

5 The communication part 110 performs communication pertaining to output of learning data, compressed learning data, and information relating to a learned regression model, as well as input of time series data or the like. The communication part 110 may be realized by a network interface card (NIC) or the like. The communication part 110 can be connected to a network by wire or wirelessly as needed, and is capable of bidirectional transmission
10 and reception of data.

[0027] (Storage Part 120)

The storage part 120 may be realized by, for example, a semiconductor memory element such as random access memory (RAM) or flash memory, or a storage device such as a hard disk or optical disk. The storage part 120 stores data and programs used in various
15 processings by the control part 130. As shown in FIG. 3, the storage part 120 includes a time series data database (DB) 121, a learning data DB 122, and a compressed learning data DB 123.

[0028] (Time Series Data DB 121)

The time series data DB 121 is a database that stores time series data collected by an external
20 data collection device such as a sensor. Here, an example of the time series data stored by the time series data DB 121 is described using FIG. 4. FIG. 4 is a table diagram showing an example of time series data according to the present embodiment.

[0029] The time series data DB 121 stores “No” being information that identifies individual data tables included in the time series data, in association with names of the
25 acquired time series data, in table format or the like. For example, as shown in FIG. 4, the

time series data DB 121 stores time series data A “a” identified by No “1”.

[0030] “a” shown in the item of the time series data A identified by No “1” is data included in a target item, and may be information such as a numerical value, text, or symbol.

As shown in FIG. 4, the time series data DB 121 is able to store multiple pieces of time series data, such as time series data A, time series data B, and time series data C.

[0031] (Learning Data DB 122)

The learning data DB 122 is a database that stores learning data extracted, from time series data, as data that satisfies a predetermined condition. Here, an example of the learning data stored by the learning data DB 122 is described using FIG. 5. FIG. 5 is a table diagram

showing an example of learning data according to the present embodiment.

[0032] The learning data DB 122 stores “No” being information for identifying individual data included in the learning data, in association with explanatory variables and objective variables, in table format or the like. For example, as shown in FIG. 5, the learning data DB 122 stores an explanatory variable (time series data A) “a” and an objective variable (time series data B) “d” identified by No “1”.

[0033] The above-mentioned “a” shown in the item of the time series data A identified by No “1” and “b” shown in the item of the time series data B are examples of data included in a target item, and may be information such as a numerical value, text, or symbol.

[0034] As shown in (1) of FIG. 5, the learning data stored by the learning data DB 122 may be expressed in a graph format with a first axis (vertical axis) as the “objective variable” and a second axis (horizontal axis) as the “explanatory variable”. In the case where the learning data includes multivariate data including 3 or more explanatory variables or objective variables, the learning data may be expressed by a graph corresponding to the number of dimensions.

[0035] (Compressed Learning Data DB 123)

The compressed learning data DB 123 is a database that stores compressed learning data which is learning data compressed by a compression part 135 to be described later. Here, an example of the compressed data stored by the compressed learning data DB 123 is described using FIG. 6. FIG. 6 is a table diagram showing an example of compressed learning data according to the present embodiment.

[0036] The compressed learning data DB 123 stores items “subspace” being information for identifying subspaces constructed by a division part 134 to be described later, in association with objective variables and explanatory variables, in table format or the like. For example, as shown in FIG. 6, the compressed learning data DB 123 stores an explanatory variable (time series data A) “A” and an objective variable (time series data B) “D” identified by subspace “1”.

[0037] The above-mentioned “A” shown in the item of the time series data A identified by subspace “1” and “B” shown in the item of the time series data B are examples of data included in a target item, and may be information such as a numerical value, text, or symbol.

[0038] As shown in (1) of FIG. 6, the compressed learning data stored by the compressed learning data DB 123 may be expressed in a graph format with a first axis (vertical axis) as the “objective variable” and a second axis (horizontal axis) as the “explanatory variable”. In the case where the compressed learning data includes multivariate data including 2 or more explanatory variables, the compressed learning data may be expressed by a graph corresponding to the number of dimensions ((2) of FIG. 6).

[0039] (Control Part 130)

The control part 130 may be realized by a processor, micro processing unit (MPU), central processing unit (CPU) or the like executing various programs stored in the storage part 120 using a RAM as a work area. The control part 130 may be realized by an integrated circuit (IC) such as an application specific integrated circuit (ASIC) or field programmable gate

array (FPGA). As shown in FIG. 3, the control part 130 includes a reception part 131, an acquisition part 132, a preprocessing part 133, the division part 134, the compression part 135, a learning part 136, and an output part 137.

[0040] (Reception Part 131)

5 The reception part 131 receives time series data input from a sensor or a field device or the like via the above-mentioned input part (not shown). The reception part 131 stores the received time series data in the storage part 120 (time series data DB 121).

[0041] The reception part 131 may automatically receive data periodically or according to a preset condition. The above-mentioned “periodically” means, for example, at
10 predetermined time intervals such as “every X minutes”, “every X hours”, or “every X days”. The phrase “preset condition” includes a condition such as “receiving time series data between 8 AM and 8 PM every day”, and is not limited if it is a condition that may be used to determine whether to execute the reception of time series data.

[0042] (Acquisition Part 132)

15 The acquisition part 132 acquires, as learning data, data that satisfies a predetermined condition, from the time series data stored in the time series data DB 121. For example, the acquisition part 132 acquires, as learning data, data serving as target objective variables and explanatory variables, from the time series data. The acquisition part 132 stores the acquired learning data in the learning data DB 122.

20 **[0043]** The acquisition part 132 may partially exclude data or combine data related to multiple periods when acquiring learning data. For example, the acquisition part 132 is able to acquire, as learning data, data of a target period based on a preset acquisition target period or exclusion target period, from the time series data.

[0044] In the present embodiment, the acquisition part 132 is described as acquiring
25 learning data from the time series data collected by a sensor installed in a chemical plant or

the like. However, the data to be acquired is not limited to time series data. For example, the acquisition part 132 is able to acquire learning data from data that can be used as learning data for constructing a regression model, such as property data for each production lot.

[0045] The acquisition part 132 is able to acquire learning data not only from sensor data but also from virtual sensor values calculated based on a predetermined model or from data that has undergone moving average processing or the like.

[0046] (Preprocessing Part 133)

The preprocessing part 133 performs predetermined preprocessing on learning data in the case where a preset condition is satisfied or where a command to execute preprocessing is issued by the user. Specifically, the preprocessing part 133 performs preprocessing on the learning data acquired by the acquisition part 132, using at least one of noise reduction, dimension compression, variable transformation, and time delay correction.

[0047] For example, the preprocessing part 133 performs noise removal on the acquired learning data by calculating a moving average. The preprocessing part 133 performs dimension compression on the acquired learning data using principal component analysis or an autoencoder.

[0048] The preprocessing part 133 executes normalization or discretization or the like and performs variable transformation on the acquired learning data. The preprocessing part 133 executes generation of a lag variable or the like and transforms the acquired learning data into data considering time delays.

[0049] Any condition such as the presence or absence of noise in data, whether the data size exceeds a specified size, or whether time delays have occurred, may be set as the above-mentioned preset condition.

[0050] (Division Part 134)

Based on an explanatory variable included in the learning data acquired by the acquisition part 132, the division part 134 divides the learning data and constructs a subspace. Specifically, the division part 134 equally divides a data section determined by a minimum value and a maximum value of the explanatory variable included in the learning data or an upper limit and a lower limit of a device from which the learning data is acquired, and constructs a subspace.

[0051] For example, the division part 134 is able to equally divide a data section between the maximum value and minimum value of each explanatory variable included in the acquired learning data, and construct a subspace. The division part 134 is able to equally divide a data section between the maximum value and minimum value in the specifications of a measuring instrument that collects time series data or the like, and construct a subspace.

[0052] Furthermore, the division part 134 divides the learning data according to the density of the learning data and constructs a subspace. For example, the division part 134 is able to construct a subspace such that the number of data points included in each explanatory variable included in the acquired learning data is equal in each subspace.

[0053] (Compression Part 135)

The compression part 135 compresses the learning data for each subspace constructed by the division part 134 and generates compressed learning data used for learning a regression model. The compression part 135 stores the generated compressed learning data in the compressed learning data DB 123.

[0054] With respect to an objective variable for each subspace obtained by division by the division part 134, the compression part 135 calculates a value of the compressed objective variable based on a predetermined calculation method. Next, according to a predetermined calculation method used to calculate the value of the compressed objective

variable, the compression part 135 calculates a value of the compressed explanatory variable for each subspace and generates compressed learning data.

[0055] The compression part 135 may execute compression according to a method set for each explanatory variable. In the case where the explanatory variable is multidimensional, the compression part 135 is able to compress a value of an objective variable included in a subspace obtained by division along an axis of each explanatory variable.

[0056] Hereinafter, a compression method for the objective variable and the explanatory variable will be further described. First, the compression of the objective variable will be described in the following items. The compression part 135 is able to calculate the value of the compressed objective variable by using one or a combination of multiple methods described below.

[0057] For example, the compression part 135 calculates an average value of the objective variable for each subspace obtained by division by the division part 134 as the value of the compressed objective variable. The compression part 135 outputs the calculated average value as the compressed objective variable.

[0058] The compression part 135 calculates a median value of the objective variable for each subspace obtained by division by the division part 134 as the value of the compressed objective variable. The compression part 135 outputs the calculated median value as the compressed objective variable.

[0059] The compression part 135 calculates a maximum value and a minimum value of the objective variable for each subspace obtained by division by the division part 134 as the value of the compressed objective variable. The compression part 135 outputs the calculated maximum value and minimum value as the compressed objective variable.

[0060] The compression part 135 calculates, as the value of the compressed objective

variable, a sampling value (for example, a sampling value for each quartile point (quartile)) corresponding to a distribution of the objective variable for each subspace obtained by division by the division part 134. The compression part 135 outputs a sampled value as the compressed objective variable.

5 **[0061]** Next, the compression of the explanatory variable is described. For example, with respect to an explanatory variable for each subspace obtained by division by the division part 134, the compression part 135 calculates a center value of a lattice constituting the subspace as the value of the compressed explanatory variable. The compression part 135 outputs the calculated explanatory variable as the compressed explanatory variable.

10 **[0062]** With respect to the explanatory variable for each subspace obtained by division by the division part 134, the compression part 135 calculates an average value of the explanatory variable included in the subspace as the value of the compressed explanatory variable. The compression part 135 outputs the calculated average value of the explanatory variable as the compressed explanatory variable.

15 **[0063]** With respect to the explanatory variable for each subspace obtained by division by the division part 134, the compression part 135 calculates a specific value included in the subspace as the value of the compressed explanatory variable. Specifically, in the case where a maximum value, a minimum value, a quantile point or the like of the objective variable included in the subspace is selected, the compression part 135 determines an
20 original position of the objective variable as the value of the explanatory variable. The compression part 135 outputs the determined value as the compressed explanatory variable.

[0064] In the above items, the compression part 135 is able to select a compression method for the explanatory variable according to a compression method of the objective variable. For example, in the case where the value of the compressed objective variable is
25 calculated as the “average value”, the compression part 135 calculates the center value of

the lattice constituting the subspace, or the maximum value and minimum value as the compressed explanatory variable. In the case where the value of the compressed objective variable is calculated as the “maximum value and minimum value” or “quantile point”, the compression part 135 calculates a specific value (original value) included in the subspace as the value of the compressed explanatory variable.

[0065] On the other hand, in addition to the above-mentioned processing of values included in the subspace, the compression part 135 performs random sampling from the data included in a subspace obtained by division by the division part 134, and calculates an obtained sampling value as the value of the compressed objective variable. For example, the compression part 135 is able to generate compressed learning data by acquiring a fixed number or fixed percentage of data from the learning data using random sampling.

[0066] (Learning Part 136)

The learning part 136 learns a regression model using the compressed learning data compressed by the compression part 135. For example, the learning part 136 is able to learn (train) a regression model using the generated compressed learning data so that, when an output value (explanatory variable) of a first sensor is input to the regression model, the regression model outputs an output value (objective variable) of a second sensor.

[0067] (Output Part 137)

The output part 137 outputs data or the regression model obtained by the processing performed by the data compression device 100 to the user or an external device. For example, the output part 137 is able to output at least one of the learning data acquired by the acquisition part 132, the compressed learning data generated by the compression part 135, and the learned regression model learned by the learning part 136.

[0068] (Example of Processing)

Hereinafter, data compression processing realized by the data compression device 100 according to the present embodiment will be described using FIG. 7 and FIG. 8. FIG. 7 and FIG. 8 are diagrams showing an example of the data compression processing according to the present embodiment.

5 **[0069]** (Generation of Learning Data for Neural Networks)

First, “generation of learning data for neural networks” is described using FIG. 7. FIG. 7 shows an example of compressing learning data 122a for neural networks with one input and one output and generating compressed learning data 123a.

[0070] First, in the data compression device 100, the learning data 122a is acquired
10 based on a predetermined condition. Next, in the data compression device 100, the data section determined by the minimum value and maximum value of the explanatory variable is equally divided ((1) of FIG. 7).

[0071] In the data compression device 100, the average value of the objective variable included in each subspace obtained by division is calculated (for example, (2-1) of FIG. 7).
15 Next, in the data compression device 100, based on the compression method for the explanatory variable corresponding to the compression method (average value) of the objective variable, the center value of the lattice of the subspace is specified (for example, (2-2) of FIG. 7). In the data compression device 100, the compressed learning data 123a is generated in which the average value of the objective variable included in the subspace
20 and the center value of the lattice are taken as the compressed explanatory variable.

[0072] FIG. 7 describes an example of one input and one output. However, in the data compression device 100, compressed learning data can be similarly generated even for learning data with multiple inputs and multiple outputs.

[0073] (Generation of Learning Data for Fuzzy Neural Networks)

Next, “generation of learning data for fuzzy neural networks” is described using FIG. 8. FIG. 8 shows an example of compressing learning data 122b for fuzzy neural networks with one input and one output and generating compressed learning data 123b.

[0074] First, as a premise, in a general regression model, there are cases where an estimated value is determined as a single point. On the other hand, in a fuzzy neural network, a range of estimated values can be obtained according to a given confidence level (for example, see Reference 1).

[0075] (Reference 1): Hisao Ishibuchi, Koichi Morioka, Hideo Tanaka (1995) “Learning of Trapezoid Fuzzy Number Weights in Fuzzy Neural Networks”, Journal of Japan Society for Fuzzy Theory and Systems, Vol. 7, No. 2, pp. 347-360 <URL: https://www.jstage.jst.go.jp/article/jfuzzy/7/2/7_KJ00002088402/_pdf/-char/ja>, <Searched on June 20, 2024>

[0076] In a fuzzy neural network, in addition to explanatory variables and objective variables, a confidence level value may be used as learning data. The confidence level is a variable that takes values in the range of 0 to 1, where 0 represents an outermost range of the estimated value, and 1 indicates an innermost range of the estimated value. For example, in the graph shown in (2) of FIG. 2, the dotted lines indicate “confidence level $h=0.0$ (upper limit)” and “confidence level $h=0.0$ (lower limit)”. In the graph shown in (2) of FIG. 2, the solid lines indicate “confidence level $h=1.0$ (upper limit)” and “confidence level $h=1.0$ (lower limit)”.

[0077] Hereinafter, a series of processing for generating learning data for fuzzy neural networks will be described. First, in the data compression device 100, the learning data 122b is acquired based on a predetermined condition. Next, in the data compression device 100, the data section determined by the minimum value and maximum value of the explanatory variable is equally divided ((1) of FIG. 8).

[0078] In the data compression device 100, the minimum value (confidence level 0) and maximum value (confidence level 0) of the objective variable included in each subspace obtained by division are calculated (for example, (2-1) of FIG. 8). In the data compression device 100, the average value (confidence level 1) of the objective variable included in each subspace obtained by division is calculated (for example, (2-2) of FIG. 8).

[0079] Next, in the data compression device 100, based on the compression method for the explanatory variable corresponding to the compression method (minimum value (confidence level 0), maximum value (confidence level 0), average value (confidence level 1)) of the objective variable, the center value of the lattice of the subspace is specified. In the data compression device 100, the compressed learning data 123b is generated in which the minimum value (confidence level 0), maximum value (confidence level 0), average value (confidence level 1) of the objective variable included in the subspace, and the center value of the lattice are taken as the compressed explanatory variable.

[0080] FIG. 8 describes an example of one input and one output. However, in the data compression device 100, compressed learning data can be similarly generated even for learning data with multiple inputs and multiple outputs.

[0081] (Processing Procedure)

Hereinafter, a processing procedure of the data compression device 100 according to the present embodiment will be described using FIG. 9. FIG. 9 is a flowchart showing an example of a procedure for data compression processing according to the present embodiment.

[0082] The reception part 131 receives time series data (S101). Next, the acquisition part 132 acquires learning data serving as a target from the time series data (S102).

[0083] Here, in the case of performing preprocessing on the learning data (S103: Yes), the preprocessing part 133 executes preprocessing of the learning data (S104). In the case

of not performing preprocessing on the learning data (S103: No), the data compression device 100 skips the process of S104.

[0084] The division part 134 divides an explanatory variable of the learning data and constructs a subspace (S105). The compression part 135 compresses the divided learning data for each subspace obtained by division and generates compressed learning data (S106).

[0085] Then, the data compression device 100 ends the process. In the data compression device 100, learning processing using the generated compressed learning data, or output processing to the user or an external device can be executed (not shown in the flowchart).

[0086] (Effects)

Hereinafter, effects achieved by the data compression device 100 according to the present embodiment will be described. If there is a biased distribution in data used to construct a general regression model such as a neural network, there are cases where the accuracy of learning is reduced. As learning data increases, more computational resources are required, and efficient learning becomes difficult.

[0087] Accordingly, the acquisition part 132 of the data compression device 100 according to the present embodiment acquires learning data serving as a target from predetermined time series data. The division part 134 of the data compression device 100 divides the learning data and constructs a subspace based on the explanatory variable included in the acquired learning data. The compression part 135 of the data compression device 100 compresses the learning data for each constructed subspace and generates compressed learning data used for learning a regression model.

[0088] As described above, in the data compression device 100, by compressing the learning data to eliminate bias in a space of the regression model, the learning data is generated that may reduce the computational resources pertaining to learning while

improving the accuracy of the regression model.

[0089] For example, in the data compression device 100, an effect is achieved that, by compressing a portion where there is bias in distribution of learning data by executing a data compression method according to the present embodiment, the computational resources
5 required for constructing the regression model are reduced.

[0090] In the data compression device 100, a problem is solved that, by eliminating bias in a regression space, a portion having high data density is selectively learned. Hence, in the data compression device 100, an effect is achieved that accuracy of the regression model is improved.

10 **[0091]** Accordingly, according to the data compression device 100 of the present embodiment produces, an effect is achieved that accurate model learning is facilitated.

[0092] Specifically, the division part 134 equally divides the data section determined by the minimum value and maximum value of the explanatory variable included in the learning data, or the upper limit and lower limit of the device from which the learning data
15 is acquired, and constructs a subspace. The division part 134 divides the learning data according to the density of the learning data, and constructs a subspace.

[0093] Through the above-mentioned processing, in the data compression device 100, based on an optimal method for constructing a subspace according to input time series data, a subspace corresponding to the number of dimensions can be constructed. In the data
20 compression device 100, by constructing the subspace corresponding to the number of dimensions, it is possible to appropriately compress various types of time series data having a biased distribution. Accordingly, in the data compression device 100, it is possible to appropriately execute the generation of compressed learning data according to the input time series data.

25 **[0094]** With respect to the objective variable for each subspace obtained by division by

the division part 134, the compression part 135 calculates the value of the compressed objective variable based on a predetermined calculation method. For example, the compression part 135 calculates at least one of the average value, median value, maximum value and minimum value, and sampling value corresponding to a distribution of the objective variable for each subspace obtained by division by the division part 134 as the value of the compressed objective variable. According to the predetermined calculation method used to calculate the value of the compressed objective variable, the compression part 135 calculates the value of the compressed explanatory variable for each subspace, and generates compressed learning data.

[0095] Through the above-mentioned processing, in the data compression device 100, the objective variable can be compressed using an optimal compression method for the objective variable corresponding to the type of acquired learning data and or purpose of learning. Hence, in the data compression device 100, by appropriately compressing various types of learning data or appropriately compressing learning data according to the purpose of learning, data used for learning for more efficient learning than conventional technology can be generated.

[0096] The compression part 135 calculates the value of the compressed explanatory variable for each subspace according to a predetermined calculation method used to calculate the value of the compressed objective variable. For example, with respect to the explanatory variable for each subspace obtained by division by the division part 134, the compression part 135 calculates at least one of the center value of the lattice constituting the subspace, the average value of the explanatory variables included in the subspace, and a specific value included in the subspace as the value of the compressed explanatory variable. The compression part 135 performs random sampling from the data included in the subspace obtained by division by the division part 134, and calculates the obtained

sampling value as the value of the compressed objective variable.

[0097] Through the above-mentioned processing, in the data compression device 100, based on an optimal compression method for the explanatory variable corresponding to the compression method for the objective variable, the explanatory variable can be compressed.

5 Hence, in the data compression device 100, by appropriately compressing various types of learning data or appropriately compressing learning data according to the purpose of learning, data used for learning for more efficient learning than conventional technology can be generated.

[0098] The preprocessing part 133 performs preprocessing on the learning data acquired
10 by the acquisition part 132, using at least one of noise reduction, dimension compression, variable transformation, and time delay correction.

[0099] As described above, in the data compression device 100, preprocessing for efficient execution of compression processing according to input time series data is performed. As a result, in the data compression device 100, an effect is achieved that
15 appropriate generation of compressed learning data is enabled by performing compression processing using the learning data that has undergone preprocessing.

[0100] The learning part 136 learns a regression model using the compressed learning data compressed by the compression part 135. Hence, in the data compression device 100, an effect is achieved that the regression model can be improved in learning accuracy and
20 can be accelerated by performing learning using the generated compressed learning data.

[0101] <Modifications>

The following describes modifications realized by the data compression device 100 according to the present embodiment.

[0102] (Data, or the Like)

The time series data, learning data, compressed learning data, regression model, subspace, names of functional parts of the data compression device 100, steps, processes, and names of steps or processes used in the description of the above embodiments are merely examples and can be arbitrarily changed.

5 **[0103]** While it has been described that the storage part 120 of the data compression device 100 according to the present embodiment stores target data or information in table format, the disclosure is not limited thereto. For example, the storage part 120 is able to store target data or information in predetermined file format, image format, or the like, in addition to table format.

10 **[0104]** For example, the time series data DB 121 stores “No” being information that identifies data acquired in time series included in the time series data as well as the name of the acquired time series data. However, the stored items and information within the items are not limited.

15 **[0105]** For example, the learning data DB 122 stores “No” being information for identifying individual data included in the learning data in association with objective variables and explanatory variables in table format or the like. However, the stored items and information within the items are not limited.

[0106] (Application Field, or the Like)

20 The data compression device 100 according to the present embodiment can be applied to various fields such as chemical plants, chemical processes, power plant facilities, machining processes (presses, numerically controlled (NC) machine tools/machining center (MCs), or the like), semiconductor manufacturing processes, and water treatment facilities. However, the above-mentioned application fields are merely examples, and the application fields of the data compression device 100 according to the present embodiment are thus not
25 limited thereto.

[0107] (Regarding Regression Model)

In the present embodiment, the regression model may be constructed based on polynomial approximation using the least squares method, neural network, fuzzy neural network, Gaussian process regression, random forest, or the like. However, the above-mentioned regression model construction methods are merely examples, and the regression model construction method of the data compression device 100 according to the present embodiment is not limited thereto.

[0108] (Flowchart, or the Like)

The steps in the flowcharts or the like may be rearranged and performed as long as there are no contradictions, or there may be steps that are not performed. Conjunctions such as “next”, “subsequently”, “furthermore”, “at this time”, and “on this occasion” in the description of the flowcharts do not limit the order or timing of the implementation of processings in the flowcharts.

[0109] (Others)

Among the processings described in the above-mentioned embodiments and modifications, all or some of the processings described as being automatically performed can also be manually performed, or all or some of the processings described as being manually performed can also be automatically performed by known methods. The processing procedures, specific names, and information including various data or parameters shown in the above document or drawings can be arbitrarily changed unless otherwise specified. For example, the various information shown in each drawing is not limited to the information shown in the drawings.

[0110] Each component of each device shown in the drawings is a functional concept and does not necessarily need to be physically configured as shown in the drawings. That is, the specific form of distribution or integration of each device is not limited to that shown

in the drawings, and all or some of the components can be configured to be functionally or physically distributed or integrated in arbitrary units according to various loads, usage conditions, or the like.

[0111] The above-mentioned components include those that can be easily assumed by those skilled in the art, those that are substantially identical, and those within the so-called range of equivalents. Furthermore, the above-mentioned embodiments and modifications can be appropriately combined within a range that does not contradict the processing contents.

[0112] The “part (section, module, unit)” as used above can be read as “means” or “circuit”, or the like. For example, a control part can be read as a control means or control circuit.

[0113] While several embodiments have been described in detail based on the drawings, these are examples, and the present embodiment can be implemented in other forms with various modifications and improvements applied therein based on the knowledge of those skilled in the art, including the aspects described in the disclosure of the invention.

[0114] <Hardware Configuration>

The data compression device 100 according to the present embodiment may be realized by, for example, a computer 1000 of a configuration as shown in FIG. 10. FIG. 10 is a hardware configuration diagram showing an example of a computer that realizes the functions of the data compression device 100 according to the present embodiment. The computer 1000 is in a form in which a CPU 1100, a RAM 1200, a ROM 1300, an auxiliary storage device 1400, a communication interface (I/F) 1500, and an input/output I/F 1600 are connected via a bus 1800.

[0115] The CPU 1100 operates based on a program stored in the ROM 1300 or the auxiliary storage device 1400 and controls each part. The ROM 1300 stores a boot

program executed by the CPU 1100 at the startup of the computer 1000 or a program dependent on the hardware of the computer 1000.

[0116] The auxiliary storage device 1400 stores a program executed by the CPU 1100 and data used by the program. The communication I/F 1500 receives data from other devices via a predetermined communication network NW (including wireless communication in a closed area in the present embodiment) and transmits the data to the CPU 1100, and transmits data generated by the CPU 1100 to other devices via the predetermined communication network NW. The CPU 1100 controls an output device such as a display or printer, and an input/output device 1700 such as a keyboard and mouse via the input/output I/F 1600. The CPU 1100 acquires data from the input/output devices 1700 via the input/output I/F 1600. The CPU 1100 outputs generated data to the input/output devices 1700 via the input/output I/F 1600.

[0117] For example, in the case where the computer 1000 functions as various devices according to the present embodiment, the CPU 1100 of the computer 1000 realizes the function of the control part by executing a program loaded on the RAM 1200.

Description of Reference Numerals

100: data compression device

110: communication part

120: storage part

121: time series data DB

122: learning data DB

123: compressed learning data DB

130: control part

131: reception part

132: acquisition part

133: preprocessing part

134: division part

135: compression part

136: learning part

5 137: output part

CLAIMS

Claim 1:

A data compression device, comprising:

5 an acquisition part, acquiring learning data serving as a target from predetermined time series data;

 a division part, dividing the learning data acquired by the acquisition part based on an explanatory variable included in the learning data, and constructing a subspace; and

 a compression part, compressing the learning data for each of the subspace
10 constructed by the division part, and generating compressed learning data used for learning a regression model.

Claim 2:

The data compression device according to claim 1, wherein

15 the division part executes at least one of:

 equally dividing a data section determined by a minimum value and a maximum value of the explanatory variable included in the learning data or by an upper limit and a lower limit of a device from which the learning data is acquired, and constructing the subspace; and

20 dividing the learning data according to a density of the learning data, and constructing the subspace.

Claim 3:

The data compression device according to claim 1 or 2, wherein

25 the compression part:

with respect to an objective variable for each of the subspace obtained by division by the division part, calculates a value of a compressed objective variable based on a predetermined calculation method;

calculates a value of a compressed explanatory variable for each of the subspace
5 according to the predetermined calculation method used to calculate the value of the compressed objective variable; and
generating the compressed learning data.

Claim 4:

10 The data compression device according to claim 3, wherein
with respect to an objective variable for each of the subspace obtained by division by the division part, the compression part calculates at least one of an average value, a median value, a maximum value and a minimum value, and a sampling value corresponding to a distribution as the value of the compressed objective variable.

15

Claim 5:

The data compression device according to claim 3, wherein
with respect to an explanatory variable of each of the subspace obtained by division by the division part, the compression part calculates at least one of a center value of a lattice
20 constituting the subspace, an average value of the explanatory variables included in the subspace, and a specific value included in the subspace as the value of the compressed explanatory variable.

Claim 6:

25 The data compression device according to claim 1 or 2, wherein

the compression part:

performs random sampling from data included in the subspace obtained by division by the division part; and

calculates an obtained sampling value as the value of the compressed objective
5 variable.

Claim 7:

The data compression device according to claim 1 or 2, further comprising:

a preprocessing part, performing preprocessing on the learning data acquired by the
10 acquisition part, using at least one of noise reduction, dimension compression, variable transformation, and time delay correction.

Claim 8:

The data compression device according to claim 1 or 2, further comprising:

15 a learning part, learning the regression model using the compressed learning data compressed by the compression part.

Claim 9:

A data compression method, executed by a data compression device, comprising:

20 an acquisition process for acquiring learning data from predetermined time series data;

a division process for dividing the learning data acquired by the acquisition process based on an explanatory variable included in the learning data, and constructing a subspace; and

a compression process for compressing the learning data for each of the subspace constructed by the division process, and generating compressed learning data used for learning a regression model.

5 Claim 10:

A data compression program, for causing a computer to execute:

an acquisition procedure, for acquiring learning data from predetermined time series data;

10 a division procedure, for dividing the learning data acquired by the acquisition procedure based on an explanatory variable included in the learning data, and constructing a subspace; and

a compression procedure, for compressing the learning data for each of the subspace constructed by the division procedure, and generating compressed learning data used for learning a regression model.