

DESCRIPTION

Title of Invention

5 PREDICTION FUNCTION CREATION DEVICE, PREDICTION
FUNCTION CREATION METHOD, AND COMPUTER-READABLE
STORAGE MEDIUM

Technical Field

10

[0001]

The present invention relates to a prediction function creation
device and a prediction function creation method for creating a prediction
function that derives an objective variable from a set of samples including
15 explanatory variables and the objective variable, and a computer-readable
recording medium recording a program for achieving the prediction
function creation device and the prediction function creation method.

Background Art

20

[0002]

In various business fields, predicting the future from obtainable
data is effective for business improvement. For example, in a store, when
future sales can be predicted from sales data during the most recent two
25 weeks, inventory control can be appropriately carried out. When it can
be predicted whether complaints come from customers and which manner
of reception operations causes such complaints from records of reception
operations at a call center, complaints can be reduced.

[0003]

30

In the description of the present invention, a type of data used as a

clue for prediction, such as sales data during the most recent two weeks and records of reception operations at a call center, is referred to as "explanatory variable", and a variable to be predicted, such as future sales and occurrence/non-occurrence of incoming complaints, is referred to as

5 "objective variable". It is assumed that "prediction" is to create a function of explanatory variables and obtain a predicted value of the objective variable.

[0004]

It is also assumed that past data are available as a clue for the

10 prediction. Past data are a set of samples each of which is a tuple of explanatory variables and an objective variable. Hereinafter, the set of samples is referred to as "training data".

[0005]

Methods to carry out prediction by use of training data include a

15 method using machine learning. Machine learning is to create a function to output a predicted value of the objective variable by using explanatory variables as input on the basis of training data.

[0006]

However, there is a problem in applying machine learning. The

20 problem is that machine learning is not applicable when an explanatory variable in the training data has a missing value. For example, when a specific item is out of stock during a certain period of time, the sale of the specific item becomes missing value, which makes machine learning inapplicable. When a portion of records of reception operations has

25 missing values because an operator who has answered a call has missed recording his/her operation, machine learning also becomes inapplicable. That is, many methods using machine learning have a problem in that the methods are not applicable to data including missing values.

[0007]

On the other hand, a method to impute missing values by a mean value and a method to impute missing values by predicting the missing value from other explanatory variables have been known. However, when a large error occurs in the imputation, these methods cause an unnecessary error in the prediction of the objective variable.

[0008]

To solve such a problem, NPL 1, for example, discloses a prediction system that can carry out prediction even when a portion of training data, which are used as input, includes missing values for the explanatory variables. Fig. 11 is a block diagram illustrating an example of a conventional prediction system.

[0009]

As illustrated in Fig. 11, a prediction system 20 includes a data partitioning means 21 and a prediction function learning means 22. When training data is input, the data partitioning means 21 partitions the input training data, and outputs the partitioned training data. When the partitioned training data is input, the prediction function learning means 22 carries out learning to create a prediction function for each partition of training data, and outputs the created prediction functions.

[0010]

The conventional prediction system illustrated in Fig. 11 operates in the following manner. First, for respective samples in the input training data, the data partitioning means 21 refers to which explanatory variable has a missing value (hereinafter, referred to as "missing manner"), and gives the same label to samples that have the same missing manner.

[0011]

Next, the prediction function learning means 22 is inputted the labeled training data output by the data partitioning means 21, carries out machine learning with respect to each label, using only a set of samples to

which the same label is given as training data, and, consequently, outputs prediction functions.

[Citation List]

5 [Non Patent Literature]

[0012]

[NPL 1] "Handling Missing Values when Applying Classification Models", Maytal Saar-Tsechansky and Foster Provost, Journal of Machine Learning Research 8 (2007) 1625-1657

10

Summary

[Technical Problem]

[0013]

15 However, there is a problem in the prediction system disclosed in the above-described NPL 1. The problem is that, when a lot of variations of missing manner in each sample are involved, accuracy in predicted values is reduced. That is because, when a lot of variations of missing manner are involved, the number of varieties of label that the data
20 partitioning means 21 gives increases in accordance. Thus, the sample size of a set of samples with the same label becomes small. When the sample size is small, accuracy in predicted values created by the prediction functions becomes low.

[0014]

25 Accordingly, even when a lot of variations of missing manner for the explanatory variables are involved, it is desirable to achieve a prediction device that can carry out prediction with high accuracy on the basis of input including missing values.

[0015]

Various embodiments provide a prediction function creation device,
a prediction function creation method, and a computer-readable recording
medium that are capable of solving the above-described problem and, even
5 when a lot of variations of missing manner for the explanatory variables
are involved, carrying out prediction with high accuracy.

10 [0016]

For achieving the above-mentioned objection, a prediction function
creation device according to an aspect for, by using a set of samples that
include explanatory variables and an objective variable as training data,
creating a prediction function to derive the objective variable, the
15 prediction function creation device includes:

a clustering unit that, in accordance with patterns of missing values
for the explanatory variables in respective samples, carries out clustering
to give labels to the respective samples, and, to each label, assigns weights
in accordance with patterns of missing values for the explanatory variables
20 in samples to which the label is given;

a child model creation unit that, on the basis of the assigned
weights, makes portions of the training data partial training data, and, on
the basis of patterns of missing values for the explanatory variables in the
samples included in the partial training data, determines an explanatory
25 variable that constitutes the prediction function among the explanatory
variables with respect to each label; and

a mixture model creation unit that, by using the explanatory
variable and the partial training data that are determined with respect to

each label, creates the prediction function with respect to each pattern of missing values.

[0017]

Furthermore, for achieving the above-mentioned objection, a
 5 prediction function creation method according to an aspect by using a set of samples that include explanatory variables and an objective variable as training data, creates a prediction function to derive the objective variable. The method includes:

(a) a step of, in accordance with patterns of missing values for the
 10 explanatory variables in respective samples, carrying out clustering to give labels to the respective samples, and, to each label, assigning weights in accordance with patterns of missing values for the explanatory variables in samples to which the label is given;

(b) a step of, on the basis of the assigned weights, making portions
 15 of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the partial training data, determining an explanatory variable that constitutes the prediction function among the explanatory variables with respect to each label; and

(c) a step of, by using the explanatory variable and the partial
 20 training data that are determined with respect to each label, creating the prediction function with respect to each pattern of missing values.

[0018]

Furthermore, for achieving the above-mentioned objection, a
 25 computer-readable recording medium according to an aspect recording a program for, by using a set of samples that include explanatory variables and an objective variable as training data, creating a prediction function to derive the objective variable, with a computer.

The program includes instructions to make the computer execute:

(a) a step of, in accordance with patterns of missing values for the explanatory variables in respective samples, carrying out clustering to give labels to the respective samples, and, to each label, assigning weights in accordance with patterns of missing values for the explanatory variables in
5 samples to which the label is given;

(b) a step of, on the basis of the assigned weights, making portions of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the partial training data, determining an explanatory variable that constitutes
10 the prediction function among the explanatory variables with respect to each label; and

(c) a step of, by using the explanatory variable and the partial training data that are determined with respect to each label, creating the prediction function with respect to each pattern of missing values.

15

[Advantageous Effects]

[0019]

As described above, with the various embodiments, it is possible to
20 carry out prediction with high accuracy even when a lot of variations of missing manner for the explanatory variables, that is, patterns of missing values for the explanatory variables, are involved.

Brief Description of Drawings

25

[0020]

[Fig. 1] Fig. 1 is a block diagram illustrating a schematic configuration of a prediction function creation device in an exemplary embodiment of the present invention;

[Fig. 2] Fig. 2 is a block diagram further specifically illustrating the configuration of the prediction function creation device illustrated in Fig. 1;

[Fig. 3] Fig. 3 is a flowchart illustrating an operation of the prediction function creation device in the exemplary embodiment of the present invention;

[Fig. 4] Fig. 4 is a block diagram illustrating an example of a computer that achieves the prediction function creation device in the exemplary embodiment of the present invention;

[Fig. 5] Fig. 5 is a diagram illustrating an example of training data that are used in an example;

[Fig. 6] Fig. 6 is a diagram illustrating an example of a result of calculation with respect to each variety of set of non-missing variables that are supposed to exist in the example;

[Fig. 7] Fig. 7 is a diagram illustrating an example of a result of calculation of distances between patterns of missing variables and candidates in the example;

[Fig. 8] Fig. 8 is a diagram illustrating an example of weights that are assigned to respective labels in the example;

[Fig. 9] Fig. 9 is a diagram illustrating an example of partial training data in the example;

[Fig. 10] Fig. 10 is a diagram illustrating a specific example of final prediction functions that are obtained in the example; and

[Fig. 11] Fig. 11 is a block diagram illustrating an example of a conventional prediction system.

Description of Embodiments

[0021]

(Summary of Invention)

The present invention has an object to prevent an increase of error when imputing explanatory variables (missing values) that are difficult to be predicted, as with the system disclosed in the above-described NPL 1.

5 In the present invention, different prediction functions are also created in accordance with patterns of missing values of the explanatory variables, as with the system disclosed in the above-described NPL 1. Since an optimum prediction function can be used with respect to each sample, it is possible to suppress a decrease in prediction accuracy.

10 [0022]

In the present invention, however, clustering of patterns of missing values for the explanatory variables is carried out, differing from the system disclosed in the above-described NPL 1. Since the number of samples that are used for each prediction function can be increased with
15 this feature, it is possible to suppress a decrease in prediction accuracy even when the number of varieties of pattern of missing values becomes large.

[0023]

In the description of the present invention, "sample" means a piece
20 of data that includes an explanatory variable and an objective variable. In general, a plurality of explanatory variables are included in a sample. The explanatory variables constitute an explanatory variable vector. The "explanatory variable vector" is a vector that has a numerical value or a sign indicating a missing value as an element.

25 [0024]

In the description of the present invention, "training data" means a set of samples. A "child model" means a group of prediction functions that a prediction function creation device of the present invention intermediately outputs. "Partial training data" means data made from a

subset of samples and variables in the training data. The partial training data is an input in learning a child model. A "final prediction function" means a prediction function to obtain values of the objective variable for respective patterns of missing values for the explanatory variables. The
5 final prediction function corresponds to an output of the prediction function creation device of the present invention.

[0025]

(Exemplary Embodiment)

Hereinafter, a prediction function creation device, a prediction
10 function creation method, and a program according to a first exemplary embodiment of the present invention will be described with reference to Figs. 1 to 4.

[0026]

[Configuration of Device]

15 First, a schematic configuration of a prediction function creation device 10 according to the exemplary embodiment will be described by using Fig. 1. Fig. 1 is a block diagram illustrating a schematic configuration of the prediction function creation device according to the exemplary embodiment of the present invention.

20 [0027]

The prediction function creation device 10, illustrated in Fig. 1, according to the exemplary embodiment is a device that creates prediction functions that derive objective variables, by using a set of samples including explanatory variables and the objective variable as training data.
25 As illustrated in Fig. 1, the prediction function creation device 10 includes a clustering unit 11, a child model creation unit 12, and a mixture model creation unit 13.

[0028]

The clustering unit 11 carries out clustering to give labels to the

respective samples in accordance with patterns of missing values for the explanatory variables in the respective samples. The clustering unit 11 also assigns weights to each label in accordance with patterns of missing values for the explanatory variables in samples to which the label is given.

5 The "patterns of missing values" referred to above include a case in which a single explanatory variable has a missing value, a case in which a plurality of explanatory variables have missing values, and a case of an empty set (a case in which no explanatory variable has a missing value).
[0029]

10 The child model creation unit 12 makes a portion of the training data the partial training data on the basis of the assigned weights. On the basis of patterns of missing values for the explanatory variables in samples included in the partial training data, the child model creation unit 12 also determines an explanatory variable (hereinafter, referred to as "prediction
15 variable") that constitutes a prediction function among the explanatory variables with respect to each label.
[0030]

By using the explanatory variables (prediction variables) each of which is determined with respect to each label and the partial training data,
20 the mixture model creation unit 13 creates a prediction function (hereinafter, referred to as "final prediction function") that derives the objective variable with respect to each pattern of missing values for the explanatory variables.
[0031]

25 As described above, in the exemplary embodiment, clustering is carried out in accordance with patterns of missing values for the explanatory variables, and, by using the partial training data created from a result of the clustering, the number of samples for each pattern of missing values can be increased. Thus, with the exemplary embodiment, it is

possible to carry out prediction with high accuracy even when a lot of patterns of missing values for the explanatory variables are involved.

[0032]

Next, by using Fig. 2, a specific configuration of the prediction function creation device 10 according to the exemplary embodiment will be described. Fig. 2 is a block diagram further specifically illustrating the configuration of the prediction function creation device illustrated in Fig. 1.

[0033]

As illustrated in Fig. 2, in the exemplary embodiment, the prediction function creation device 10 is constructed on a computer 100 with a program. To the computer 100, a terminal 110 for users is connected via a network or the like. The training data are transmitted from the terminal 110. As illustrated in Fig. 2, the prediction function creation device 10 further includes a missing variable imputation unit 14, in addition to the clustering unit 11, the child model creation unit 12, and the mixture model creation unit 13 which are mentioned above.

[0034]

In the exemplary embodiment, with respect to each variety of set of explanatory variables without missing values corresponding to respective patterns of missing values for the explanatory variables, the clustering unit 11 identifies a group of samples applying to the variety, that is, a group of samples in which a set of explanatory variables constituting one of the varieties have no missing value.

[0035]

Next, with respect to each pattern of missing values for the explanatory variables, the clustering unit 11 identifies a sample that is included in common in a sample that does not apply to the pattern of missing value and the identified group of samples.

[0036]

The clustering unit 11 then gives a label to each group of the identified common samples the number of which is greater than or equal to a certain value. Thereafter, the clustering unit 11 assigns a weight to
5 each label in such a way that, the greater the number of common samples identified with respect to each pattern of missing values is, the larger the value of the weight becomes.

[0037]

In the exemplary embodiment, the child model creation unit 12
10 make a group of samples, to which labels with weights of a threshold value or greater are given, the partial training data. Next, with respect to each label with weight, the child model creation unit 12 determines a prediction variable in the partial training data to which the label is given. The prediction variable at this time is required to have features that a number
15 of samples which itself does not have a missing value increases, and a number of variables increases, for each label.

[0038]

In the exemplary embodiment, first, with respect to each label with the weight, the mixture model creation unit 13 creates a prediction
20 function (hereinafter, referred to as "intermediate prediction function") that derives the objective variable from the prediction variable by using the partial training data to which the label is given. With respect to each pattern of missing values, the mixture model creation unit 13 creates a final prediction function by using a weight assigned for the pattern and the
25 intermediate prediction function created for a label to which the weight is assigned.

[0039]

In other words, in the exemplary embodiment, the mixture model creation unit 13 learns a child model from the partial training data

extracted by the child model creation unit 12, by using a prediction variable, with respect to each label. The mixture model creation unit 13 learns a mixture model with respect to each label by adding the learned child models.

5 [0040]

When, in partial training data, there is a sample in which a prediction variable (an explanatory variable) of a label given to the partial training data has a missing value, the missing variable imputation unit 14 imputes the prediction variable that has a missing value in the sample. In
10 this case, the mixture model creation unit 13 creates an intermediate prediction function by using a set of the imputed partial training data.
[0041]

Specifically, first, with respect to each prediction variable, the missing variable imputation unit 14 learns a function that estimates the
15 prediction variable from the values of other explanatory variables as the partial training data to be input data. When a sample to be imputed is input from the child model creation unit 12, the missing variable imputation unit 14 estimates a missing value in the sample by using the learned function, and outputs the sample including the estimated value
20 (partial training data) to the mixture model creation unit 13.
[0042]

[Operation of Device]

Next, an operation of the prediction function creation device 10 according to the exemplary embodiment of the present invention will be
25 described by using Fig. 3. Fig. 3 is a flowchart illustrating an operation of the prediction function creation device according to the exemplary embodiment of the present invention. In the following description, Figs. 1 and 2 are taken into consideration appropriately. In the exemplary embodiment, operating the prediction function creation device 10 causes

the prediction function creation method to be carried out. Thus, the description of the prediction function creation method according to the exemplary embodiment will be substituted by the following description of an operation of the prediction function creation device 10.

5 [0043]

As illustrated in Fig. 3, first, the clustering unit 11 identifies patterns of missing values for the explanatory variables in respective samples that constitute input training data (step A1). Next, the clustering unit 11, in accordance with the patterns of missing values for the explanatory variables in the respective samples, carries out clustering to assign weights to each label for the respective samples (step A2).

[0044]

The "clustering" in step A2 is to assign labels and weights, numbers of which are finite, to the respective patterns of missing values.

15 [0045]

Specifically, in step A2, with respect to each variety of set of explanatory variables without missing values in respective patterns of missing values for the explanatory variables, the clustering unit 11 identifies a group of samples applying to the variety (a group of samples in which a set of explanatory variables constituting the variety have no missing value). With respect to each pattern of missing values for the explanatory variables, the clustering unit 11 identifies a sample that is included in common in a sample that does not apply to the pattern and each of the identified groups of samples. The clustering unit 11 selects candidates in which the numbers of the identified samples are greater than or equal to a certain value, and gives a label to each of the selected candidates. The clustering unit 11 assigns a weight to each label in such a way that, the greater the number of the common samples identified with respect to each pattern of missing values is, the larger the value of the

weight becomes.

[0046]

With respect to each label given in step A2, learning of a "child model", which will be described later, is performed. In the exemplary embodiment, each label that is given through the clustering in step A2 will
5 be denoted by "L" in the following description.

[0047]

The child model creation unit 12 identifies candidates to which labels L with weights are given, and extracts a group of samples included
10 in the identified candidates as the partial training data (step A3). The child model creation unit 12 also associates a weight assigned to a label L with each sample that constitutes the partial training data as the weight of the sample.

[0048]

15 Next, with respect to each label with weight, the child model creation unit 12 determines a prediction variable among the explanatory variables in the partial training data to which the label is given (step A4).

[0049]

Next, in the partial training data extracted in step A3, the missing
20 variable imputation unit 14 confirms whether or not the prediction variable, which is determined in step A4, of the label given to the partial training data has a missing value in any sample. When there is a missing value, the missing variable imputation unit 14 imputes the prediction variable having a missing value in the sample, and updates the partial training data
25 (step A5).

[0050]

Next, with respect to each label with weight, the mixture model creation unit 13 learns an intermediate prediction function (child model) that derives the objective variable from the prediction variable by using

the partial training data to which the label is given (step A6).

[0051]

Next, when the mixture model creation unit 13 has learned child models for all labels, the mixture model creation unit 13 learns a final prediction function with respect to each pattern of missing values by using a weight assigned to the pattern and a child model for the label to which the weight is assigned (step A7).

[0052]

As described above, in the exemplary embodiment, different final prediction functions are created corresponding to patterns of missing values for the explanatory variables. Among the patterns of missing values, patterns for which the common variable is used as a prediction variable are grouped together, and learning is performed with common samples. In other words, in the exemplary embodiment, with respect to each pattern of missing values, samples that have the common explanatory variable without missing value are grouped together, and learning is performed, in the creation of final prediction function. Thus, with the exemplary embodiment, it is possible to carry out prediction with high accuracy even when a lot of patterns of missing values for the explanatory variable are involved.

[0053]

[Program]

A program according to the exemplary embodiment may be a program that causes a computer to execute steps A1 to A7 illustrated in Fig. 3. By installing the program into the computer and executing the program, it is possible to achieve the prediction function creation device 10 and the prediction function creation method according to the exemplary embodiment. In this case, a CPU (Central Processing Unit) in the computer functions as the clustering unit 11, the child model creation unit

12, the mixture model creation unit 13, and the missing variable imputation unit 14, and performs the processing of them.

[0054]

The computer that achieves the prediction function creation device
5 10 by executing the program according to the exemplary embodiment will be described by using Fig. 4. Fig. 4 is a block diagram illustrating an example of the computer that achieves the prediction function creation device according to the exemplary embodiment of the present invention.

[0055]

10 As illustrated in Fig. 4, a computer 100 includes a CPU 111, a main memory 112, a storage device 113, an input interface 114, a display controller 115, a data reader/writer 116, and a communication interface 117. The each unit is connected with each other via a bus 121 in a data communicable manner.

15 [0056]

The CPU 111 deploys the program (codes), which is stored in the storage device 113, according to the exemplary embodiment into the main memory 112, and performs each type of operations by executing them in a predetermined sequence. The main memory 112 is typically a transitory
20 storage device such as a DRAM (Dynamic Random Access Memory). The program according to the exemplary embodiment is provided with a state stored in the computer-readable recording medium 120. The program according to the exemplary embodiment may be a program that is distributed on the Internet, which is connected via the communication
25 interface 117.

[0057]

Specific examples of the storage device 113 include a semiconductor storage device, such as a flash memory, besides a hard disk drive. The input interface 114 mediates data transmission between the

CPU 111 and the input device 118, such as a keyboard and a mouse. The display controller 115 is connected with a display device 119, and controls display on the display device 119.

[0058]

5 The data reader/writer 116 mediates data transmission between the CPU 111 and the recording medium 120, and performs to read the programs from the recording medium 120 and write a processing result of the computer 100 into the recording medium 120. The communication interface 117 mediates data transmission between the CPU 111 and other
10 computers.

[0059]

 Specific examples of the recording medium 120 include a generic semiconductor storage device, such as a CF (Compact Flash (Registered Trademark)) and an SD (Secure Digital), a magnetic storage device, such
15 as a flexible disk, or an optical storage medium, such as a CD-ROM (Compact Disk Read Only Memory).

 Example

[0060]

 An example corresponding to the above-described exemplary
20 embodiment will be described by using Figs. 5 to 10. The following description will be made along the respective steps illustrated in Fig. 3.

[0061]

 Fig. 5 is a diagram illustrating an example of training data used in the example. As illustrated in Fig. 5, in the example, a number is given
25 to every sample that constitutes the training data. Every sample has three variables X1, X2, and X3 as explanatory variables and a variable Y as an objective variable. Missing values for explanatory variables are denoted by sign "NA"s.

[0062]

In Fig. 5, it is assumed that a cell with "*" marked has any numerical value. In the example in Fig. 5, samples with numbers 1 to 5 have no missing value, samples with numbers 6 to 10 have missing values only for X1, samples with numbers 11 to 20 have missing values for X1 and X2, and samples with numbers 21 to 40 have missing values for X3. Specific operations when such training data are input are as follows.

[0063]

[Step A1]

First, for respective samples in the training data, the clustering unit 11 identifies assumable patterns of missing values for the explanatory variables (hereinafter, referred to as "pattern of missing variable"). Since the explanatory variables in the samples are X1, X2, and X3, the clustering unit 11 identifies eight patterns as patterns of missing values, namely, an empty set (no missing value), a missing value for X1, a missing value for X2, a missing value for X3, missing values for X1 and X2, missing values for X2 and X3, missing values for X1 and X3, and missing values for all X1 to X3.

[0064]

In the example in Fig. 5, for samples with numbers 1 to 40, the clustering unit 11 determines that samples with numbers 1 to 5 have no missing value, samples with numbers 6 to 10 have a missing value for X1, samples with numbers 11 to 20 have missing values for X1 and X2, and samples with numbers 21 to 40 have a missing value for X3.

[0065]

[Step A2]

Next, in accordance with patterns of missing values for the explanatory variables in the respective samples, the clustering unit 11 carries out clustering to give labels to the respective samples, and assigns weights to the respective labels.

[0066]

Specifically, as described below, the clustering unit 11 carries out clustering to allocate labels and weights to the samples. First, the clustering unit 11 identifies a case in which all three explanatory variables have no missing value, a case in which two explanatory variables have no missing value (only one explanatory variable has a missing value), a case in which only one explanatory variable has no missing value (two explanatory variables have missing values), a case in which all three explanatory variables have missing values, and the like. In the example, since the number of explanatory variables is three, eight patterns actually exist as the assumable cases. It is assumed that a set of explanatory variables with no missing value (hereinafter, referred to as "set of non-missing variables") in each of the assumable cases is A. The clustering unit 11 carries out the following calculation with respect to each variety of set A of non-missing variables. In the example, the calculation is carried out by multiplying [the number of samples which do not miss A] by [the number of elements in A] (= [the number of samples which do not miss A]*[the number of elements in A]).

[0067]

Fig. 6 is a diagram illustrating an example of a result of the calculation with respect to each variety of set of non-missing variables supposed in the example. When multiplied values illustrated in Fig. 6 are obtained, by making all of A(s), the multiplied values of which are greater than or equal to a threshold value, candidates, the clustering unit 11 gives, as a label, a number allocated for each candidate to the group of samples that constitutes the respective candidates.

[0068]

In the example in Fig. 6, it is assumed that the threshold value is set at 20. In this case, the clustering unit 11 makes respective sets of

non-missing variables the multiplied values of which are greater than or equal to the threshold value, that is, X_1 , X_2 , X_3 , (X_1, X_2) , and (X_2, X_3) , the candidates, and gives numbers indicating labels to them (refer to Fig. 9). In the description below, the candidates are also denoted by a

5 candidate (X_1), a candidate (X_2), a candidate (X_3), a candidate (X_1, X_2), and a candidate (X_2, X_3). Further, as illustrated in Fig. 8, which will be described later, numbers 5, 1, 2, 3, and 4 are given to the candidate (X_1), the candidate (X_2), the candidate (X_3), the candidate (X_1, X_2), and the candidate (X_2, X_3) as labels in this order.

10 [0069]

Next, with respect to each assumable pattern of missing variables, the clustering unit 11 calculates a distance $d(I, L)$ from each candidate. Specifically, the clustering unit 11 identifies an explanatory variable that has no missing value in common in both, and calculates a distance in such a

15 way that, the greater the number of samples in which the identified explanatory variable has no missing values is, the larger the value of distance d becomes. For example, since an explanatory variable that has no missing value in common is X_2 for the pattern X_1 of missing variable and the candidate (X_2), a distance d is calculated on the basis of the

20 number of samples in which the explanatory variable X_2 has no missing value.

[0070]

For example, by assuming that a set of samples that include the explanatory variable without missing value for a pattern of missing

25 variables is denoted by I_1 and a set of samples that include the explanatory variable without missing value for a candidate with a label L is denoted by I_2 , the clustering unit 11 is able to calculate a distance d by using the following equation 1.

[0071]

[Equation 1]

$$d(I, L) := |I_1 \setminus I_2| + \alpha |I_2 \setminus I_1|$$

[0072]

In the above-described equation 1, α denotes an arbitrary integer, I denotes an arbitrary pattern of missing variables, and L denotes an arbitrary label number. It is also assumed that $|\cdot|$ indicates the number of elements in a set " $\cdot$ ", and " $A \setminus B$ " indicates a set of elements included in a set A but not included in a set B. A result of calculation in the case of $\alpha=2$ in the above-described equation 1 is illustrated in Fig. 7. Fig. 7 is a diagram illustrating an example of a result of calculation of the distances between patterns of missing variables and candidates in the example.

[0073]

Next, for each pattern I of missing variables, when the distance d to each candidate is less than a threshold value, the clustering unit 11 calculates a similarity to the candidate the distance d to which is less than the threshold value by using the following equation 2.

[0074]

[Equation 2]

$$\text{Similarity} = e^{-\lambda d(I, L)}$$

[0075]

Next, the clustering unit 11 normalizes the respective similarities by using the following equation 3 in such a way that, for each pattern I of missing variables, the sum of the similarities, which are calculated for candidates the distances d to which are less than the threshold value, is 1. The clustering unit 11 outputs the obtained values as weights $w(I, L)$ of labels L with respect to patterns I of missing variables.

[0076]

[Equation 3]

$$w(I, L) := \frac{e^{-\lambda d(I, L)}}{\sum_L e^{-\lambda d(I, L)}}$$

[0077]

A specific example of the result of calculation by the above-described equation 3 is illustrated in Fig. 8. Fig. 8 is a diagram illustrating an example of weights assigned to the respective labels in the example. The weights in the example in Fig. 8 are calculated under the assumption that the threshold value for distances d is 2.

[0078]

In the example, it is also possible to carry out clustering by a method other than the above-described method in which assumable sets of non-missing variables (candidates) are used. For example, by defining a $\{0, 1\}$ value vector with dimensions equal to the number of explanatory variables in such a way that, at each sample, a value of element is set 1 when the explanatory variable of element has a missing value and a value of element is set 0 when the explanatory variable of element has no missing value, it is also possible to carry out clustering by using an existing k-means method, dendrogram method, or the like.

[0079]

[Step A3]

Next, the child model creation unit 12 extracts samples, that become patterns of missing variables with large weights assigned to labels, from the training data, associates sample weights with the extracted samples in such a way that, the larger a weight assigned to a label is, the larger a sample weight becomes, and outputs the extracted samples as partial training data.

[0080]

For example, the child model creation unit 12 identifies sets of samples with positive weights assigned to labels L with respect to each

pattern of missing variables, and associates a weight assigned to a label L with each identified sample as a sample weight. The child model creation unit 12 makes sets of samples associated with sample weights the partial training data.

5 [0081]

[Step A4]

Next, with respect to each label, the child model creation unit 12 selects an explanatory variable (a prediction variable) used for prediction among the explanatory variables. At that time, in partial training data,
10 the child model creation unit 12 increases the numbers of the explanatory variables selected as prediction variables as much as possible, and the number of samples in which the explanatory variable selected as a prediction variable has no missing value greater than the number of samples in which the other explanatory variable has no missing value.

15 [0082]

Specifically, in the example, the child model creation unit 12 makes an explanatory variable (hereinafter, referred to as "candidate variable"), which constitutes a set of non-missing variables determined as a candidate at clustering, a prediction variable for each label. Thus, the child model
20 creation unit 12 selects prediction variables, such as X2 for the label 1, X3 for the label 2, X1 and X2 for the label 3, X2 and X3 for label 4, and X1 for the label 5. Furthermore, at this time, the child model creation unit 12 identifies a sample that has no missing value for any candidate variables in the training data, and adds it to the partial training data.

25 [0083]

In the example, a prediction variable is selected from explanatory variables which have no missing value in partial training data given a label given. However, the example is not intended to this. For example, a prediction variable may be selected from explanatory variables which have

missing values in partial training data, and imputed thereafter.

[0084]

Fig. 9 is a diagram illustrating an example of partial training data according to the example. As illustrated in Fig. 9, in the example, the partial training data is identified for every label. Sample numbers illustrated in Fig. 9 coincide with the numbers of respective samples illustrated in Fig. 3.

[0085]

[Step A5]

Next, when there is a sample which has a missing value for a prediction variable in partial training data, the missing variable imputation unit 14 performs imputation for the sample. For example, the missing variable imputation unit 14 is capable of performing imputation by substituting, with a mean value of other prediction variables, a value of a prediction variable having a missing value in a sample in which the prediction variable has a missing value. The missing variable imputation unit 14 is also capable of performing imputation by using a single substitution method, a multiple substitution method, or the like, which perform regression on other variables. In the example, the partial training data is identified in such a way that no sample in which a prediction variable has a missing value is generated.

[0086]

[Step A6]

Next, with respect to each label L, the mixture model creation unit 13 creates a prediction function from the prediction variable to the objective variable by using the partial training data, for example, by using a method such as linear regression or Support Vector Regression. In this case, the mixture model creation unit 13 calculates a loss function or a likelihood function by weighting in accordance with a weight of the

sample.

[0087]

Specifically, first, by assuming that a weight of a sample i is w_i and a loss function of a model is $l(x_i, y_i)$, the mixture model creation unit 13 estimates a parameter β by using the following equation 4, and creates a prediction function f_L including β with respect to each label L .

[0088]

[Equation 4]

$$\hat{\beta} = \arg \min_{\beta} \sum_i w_i l(x_i, y_i)$$

10 [0089]

For example, when performing linear regression in such a way as to minimize a square error, the mixture model creation unit 13 estimates a parameter β by using the following equation 5, and creates a prediction function f_L including β with respect to each label L .

15 [0090]

[Equation 5]

$$\hat{\beta} = \arg \min_{\beta} \sum_i w_i |y_i - (\beta^T x_i + \beta_0)|^2$$

[0091]

[Step A7]

20 Last, the mixture model creation unit 13 creates a final prediction function $f_I(x)$ with respect to each pattern I of missing variables in such a way that, the larger the weight $w(I, L)$ assigned to a label L is, the more importance is placed on the prediction function created for the label L . A specific example of the final prediction function $f_I(x)$ is expressed by the following equation 6.

25

[0092]

[Equation 6]

$$f_I(x) := \sum_L w(I, L) f_L(x)$$

[0093]

Fig. 10 is a diagram illustrating a specific example of the final prediction functions obtained in the example. The final prediction functions illustrated in Fig. 10 are obtained by also calculating $w(I, L)$ for patterns of missing variables that are not illustrated in the example of training data in Fig. 3.

[0094]

As described above, in the example, different final prediction functions corresponding to patterns of missing variables are created, as illustrated in Fig. 10. Thus, the example shows that, even when a lot of patterns of missing values for the explanatory variables are involved, it is possible to carry out prediction with high accuracy.

[0095]

The whole or part of the exemplary embodiments and the examples disclosed above can be described as, but not limited to, the following supplementary note 1 to the supplementary note 15.

[0096]

(Supplementary note 1)

A prediction function creation device for, by using a set of samples that include explanatory variables and an objective variable as training data, creating a prediction function to derive the objective variable, the prediction function creation device includes:

a clustering unit that, in accordance with patterns of missing values for the explanatory variables in respective samples, carries out clustering to give labels to the respective samples, and, to each label, assigns weights in accordance with patterns of missing values for the explanatory variables in samples to which the label is given;

a child model creation unit that, on the basis of the assigned weights, makes portions of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the partial training data, determines an explanatory variable that constitutes the prediction function among the explanatory variables with respect to each label; and

a mixture model creation unit that, by using the explanatory variable and the partial training data that are determined with respect to each label, creates the prediction function with respect to each pattern of missing values.

[0097]

(Supplementary note 2)

The prediction function creation device according to supplementary note 1,

wherein

the clustering unit

identifies an applied group of samples with each variety of set of explanatory variables without missing values corresponding to the respective patterns of missing values,

further, with respect to each of the patterns of missing values, identifies a sample that is included in common in a samples that does not apply to the pattern and each of the identified groups of samples,

gives a label to each group of the identified common samples the number of which is greater than or equal to a certain value, and

thereafter, with respect to each of the patterns of missing values, assigns the weight to each label in such a way that, the greater the number of the identified common samples is, the larger the value of the weight becomes.

[0098]

(Supplementary note 3)

The prediction function creation device according to supplementary note 2,

wherein

5 the child model creation unit

makes groups of samples assigned labels with the weights the partial training data, and

with respect to each label with the weight, selects an explanatory variable that constitutes the prediction function among the explanatory
10 variables.

[0099]

(Supplementary note 4)

The prediction function creation device according to supplementary note 3,

15 wherein

the mixture model creation unit,

with respect to each label with the weight, creates a second prediction function that derives the objective variable from the determined explanatory variable by using the partial training data given the label is
20 given, and

with respect to each of the patterns of missing values, creates the prediction function by using the weight that is assigned for the pattern and the second prediction function for a label to which the weight is assigned.

[0100]

25 (Supplementary note 5)

The prediction function creation device according to supplementary note 4, further includes:

a missing variable imputation unit that, when a sample in which an explanatory variable, which is determined for a label given to the partial

training data and constitutes the prediction function, is missed exists in the partial training data, imputes the explanatory variable that is missed in the sample,

wherein the mixture model creation unit creates the second
 5 prediction function by using the imputed partial training data.

[0101]

(Supplementary note 6)

A prediction function creation method for, by using a set of samples that include explanatory variables and an objective variable as training
 10 data, creating a prediction function to derive the objective variable, the method includes:

(a) a step of, in accordance with patterns of missing values for the explanatory variables in respective samples, carrying out clustering to give labels to the respective samples, and, to each label, assigning weights in
 15 accordance with patterns of missing values for the explanatory variables in samples to which the label is given;

(b) a step of, on the basis of the assigned weights, making portions of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the
 20 partial training data, determining an explanatory variable that constitutes the prediction function among the explanatory variables with respect to each label; and

(c) a step of, by using the explanatory variable and the partial training data that are determined with respect to each label, creating the
 25 prediction function with respect to each pattern of missing values.

[0102]

(Supplementary note 7)

The prediction function creation method according to supplementary note 6,

wherein

in the step (a),

identifying an applied group of samples with each variety of set of explanatory variables without missing values corresponding to the

5 respective patterns of missing values,

 further, with respect to each of the patterns of missing values, identifying a sample that is included in common in a samples that does not apply to the pattern and each of the identified groups of samples,

 giving a label to each group of the identified common samples the
10 number of which is greater than or equal to a certain value, and

 thereafter, with respect to each of the patterns of missing values, assigning the weight to each label in such a way that, the greater the number of the identified common samples is, the larger the value of the weight becomes.

15 [0103]

 (Supplementary note 8)

 The prediction function creation method according to supplementary note 7,

 in the step (b),

20 making groups of samples assigned labels with the weights the partial training data, and

 with respect to each label with the weight, selecting an explanatory variable that constitutes the prediction function among the explanatory variables.

25 [0104]

 (Supplementary note 9)

 The prediction function creation method according to supplementary note 8,

 in the step (c),

the mixture model creation unit,

with respect to each label with the weight, creating a second prediction function that derives the objective variable from the determined explanatory variable by using the partial training data given the label is
5 given, and

with respect to each of the patterns of missing values, creating the prediction function by using the weight that is assigned for the pattern and the second prediction function for a label to which the weight is assigned.
[0105]

10 (Supplementary note 10)

The prediction function creation method according to supplementary note 9, further includes:

(d) a step of, when a sample in which an explanatory variable, which is determined for a label given to the partial training data and
15 constitutes the prediction function, is missed exists in the partial training data, imputing the explanatory variable that is missed in the sample,

wherein in the step (c), creating the second prediction function by using the imputed partial training data.

[0106]

20 (Supplementary note 11)

A computer-readable recording medium recording a program for, by using a set of samples that include explanatory variables and an objective variable as training data, creating a prediction function to derive the objective variable, with a computer, the program including instructions to
25 make the computer execute:

(a) a step of, in accordance with patterns of missing values for the explanatory variables in respective samples, carrying out clustering to give labels to the respective samples, and, to each label, assigning weights in accordance with patterns of missing values for the explanatory variables in

samples to which the label is given;

(b) a step of, on the basis of the assigned weights, making portions of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the partial training data, determining an explanatory variable that constitutes the prediction function among the explanatory variables with respect to each label; and

(c) a step of, by using the explanatory variable and the partial training data that are determined with respect to each label, creating the prediction function with respect to each pattern of missing values.

[0107]

(Supplementary note 12)

The computer-readable recording medium according to supplementary note 11,

wherein

in the step (a),

identifying an applied group of samples with each variety of set of explanatory variables without missing values corresponding to the respective patterns of missing values,

further, with respect to each of the patterns of missing values, identifying a sample that is included in common in a samples that does not apply to the pattern and each of the identified groups of samples,

giving a label to each group of the identified common samples the number of which is greater than or equal to a certain value, and

thereafter, with respect to each of the patterns of missing values, assigning the weight to each label in such a way that, the greater the number of the identified common samples is, the larger the value of the weight becomes.

[0108]

(Supplementary note 13)

The computer-readable recording medium according to
supplementary note 12,

in the step (b),

5 making groups of samples assigned labels with the weights the
partial training data, and

with respect to each label with the weight, selecting an explanatory
variable that constitutes the prediction function among the explanatory
variables.

10 [0109]

(Supplementary note 14)

The computer-readable recording medium according to
supplementary note 13,

in the step (c),

15 the mixture model creation unit,

with respect to each label with the weight, creating a second
prediction function that derives the objective variable from the determined
explanatory variable by using the partial training data given the label is
given, and

20 with respect to each of the patterns of missing values, creating the
prediction function by using the weight that is assigned for the pattern and
the second prediction function for a label to which the weight is assigned.

[0110]

(Supplementary note 15)

25 The computer-readable recording medium according to
supplementary note 14, further including instructions to make the computer
execute:

(d) a step of, when a sample in which an explanatory variable,
which is determined for a label given to the partial training data and

constitutes the prediction function, is missed exists in the partial training data, imputing the explanatory variable that is missed in the sample,

wherein in the step (c), creating the second prediction function by using the imputed partial training data.

5 [0111]

While the invention has been particularly shown and described with reference to exemplary embodiments thereof, the invention is not limited to these embodiments. It will be understood by those of ordinary skill in the art that various changes in form and details may be made therein without departing from the spirit and scope of the present invention as defined by the claims.

[0112]

This application is based upon and claims the benefit of priority from Japanese patent application No. 2013-123482, filed on June 12, 2013, the disclosure of which is incorporated herein in its entirety by reference.

[Industrial Applicability]

[0113]

As described above, with the present invention, it is possible to carry out prediction with high accuracy even when a lot of variations of missing manner for the explanatory variables, that is, patterns of missing values for the explanatory variables, are involved. The present invention is useful for the field of predictive analysis, such as analysis of business log and purchase prediction in the future from a purchase history.

[Reference Signs List]

[0114]

11 Clustering unit

	12 Child model creation unit
	13 Mixture model creation unit
	14 Missing variable imputation unit
	100 Computer
5	110 Terminal
	111 CPU
	112 Main memory
	113 Storage device
	114 Input interface
10	115 Display controller
	116 Data reader/writer
	117 Communication interface
	118 Input device
	119 Display device
15	120 Recording medium
	121 Bus

CLAIMS

1. A prediction function creation device for, by using a set of samples that include explanatory variables and an objective variable as training data, creating a prediction function to derive the objective variable, the prediction function creation device comprising:

a clustering unit that, in accordance with patterns of missing values for the explanatory variables in respective samples, carries out clustering to give labels to the respective samples, and, to each label, assigns weights in accordance with patterns of missing values for the explanatory variables in samples to which the label is given;

a child model creation unit that, on the basis of the assigned weights, makes portions of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the partial training data, determines an explanatory variable that constitutes the prediction function among the explanatory variables with respect to each label; and

a mixture model creation unit that, by using the explanatory variable and the partial training data that are determined with respect to each label, creates the prediction function with respect to each pattern of missing values.

2. The prediction function creation device according to Claim 1, wherein

the clustering unit

identifies an applied group of samples with each variety of set of explanatory variables without missing values corresponding to the respective patterns of missing values,

further, with respect to each of the patterns of missing values,

identifies a sample that is included in common in a samples that does not apply to the pattern and each of the identified groups of samples,

gives a label to each group of the identified common samples the number of which is greater than or equal to a certain value, and

5 thereafter, with respect to each of the patterns of missing values, assigns the weight to each label in such a way that, the greater the number of the identified common samples is, the larger the value of the weight becomes.

10 3. The prediction function creation device according to Claim 2, wherein

the child model creation unit

makes groups of samples assigned labels with the weights the partial training data, and

15 with respect to each label with the weight, selects an explanatory variable that constitutes the prediction function among the explanatory variables.

20 4. The prediction function creation device according to Claim 3, wherein

the mixture model creation unit,

with respect to each label with the weight, creates a second prediction function that derives the objective variable from the determined explanatory variable by using the partial training data given the label is
25 given, and

with respect to each of the patterns of missing values, creates the prediction function by using the weight that is assigned for the pattern and the second prediction function for a label to which the weight is assigned.

5. The prediction function creation device according to Claim 4, further comprising:

a missing variable imputation unit that, when a sample in which an explanatory variable, which is determined for a label given to the partial training data and constitutes the prediction function, is missed exists in the partial training data, imputes the explanatory variable that is missed in the sample,

wherein the mixture model creation unit creates the second prediction function by using the imputed partial training data.

10

6. A prediction function creation method for, by using a set of samples that include explanatory variables and an objective variable as training data, creating a prediction function to derive the objective variable, the method comprising:

15 (a) a step of, in accordance with patterns of missing values for the explanatory variables in respective samples, carrying out clustering to give labels to the respective samples, and, to each label, assigning weights in accordance with patterns of missing values for the explanatory variables in samples to which the label is given;

20 (b) a step of, on the basis of the assigned weights, making portions of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the partial training data, determining an explanatory variable that constitutes the prediction function among the explanatory variables with respect to
25 each label; and

(c) a step of, by using the explanatory variable and the partial training data that are determined with respect to each label, creating the prediction function with respect to each pattern of missing values.

7. A computer-readable recording medium recording a program for, by using a set of samples that include explanatory variables and an objective variable as training data, creating a prediction function to derive the objective variable, with a computer, the program comprising

5 instructions to make the computer execute:

(a) a step of, in accordance with patterns of missing values for the explanatory variables in respective samples, carrying out clustering to give labels to the respective samples, and, to each label, assigning weights in accordance with patterns of missing values for the explanatory variables in
10 samples to which the label is given;

(b) a step of, on the basis of the assigned weights, making portions of the training data partial training data, and, on the basis of patterns of missing values for the explanatory variables in the samples included in the partial training data, determining an explanatory variable that constitutes
15 the prediction function among the explanatory variables with respect to each label; and

(c) a step of, by using the explanatory variable and the partial training data that are determined with respect to each label, creating the prediction function with respect to each pattern of missing values.

Fig.1

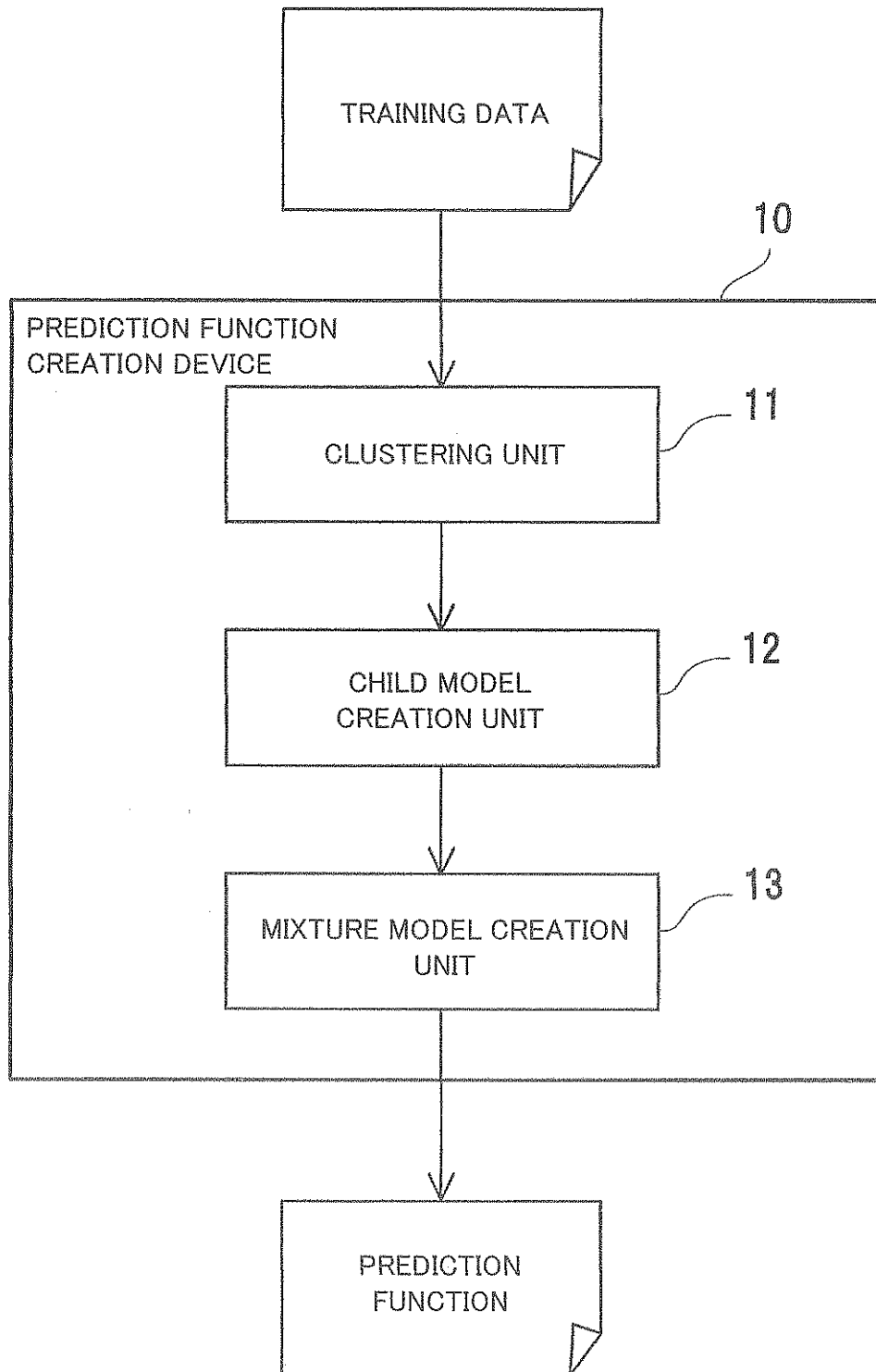


Fig.2

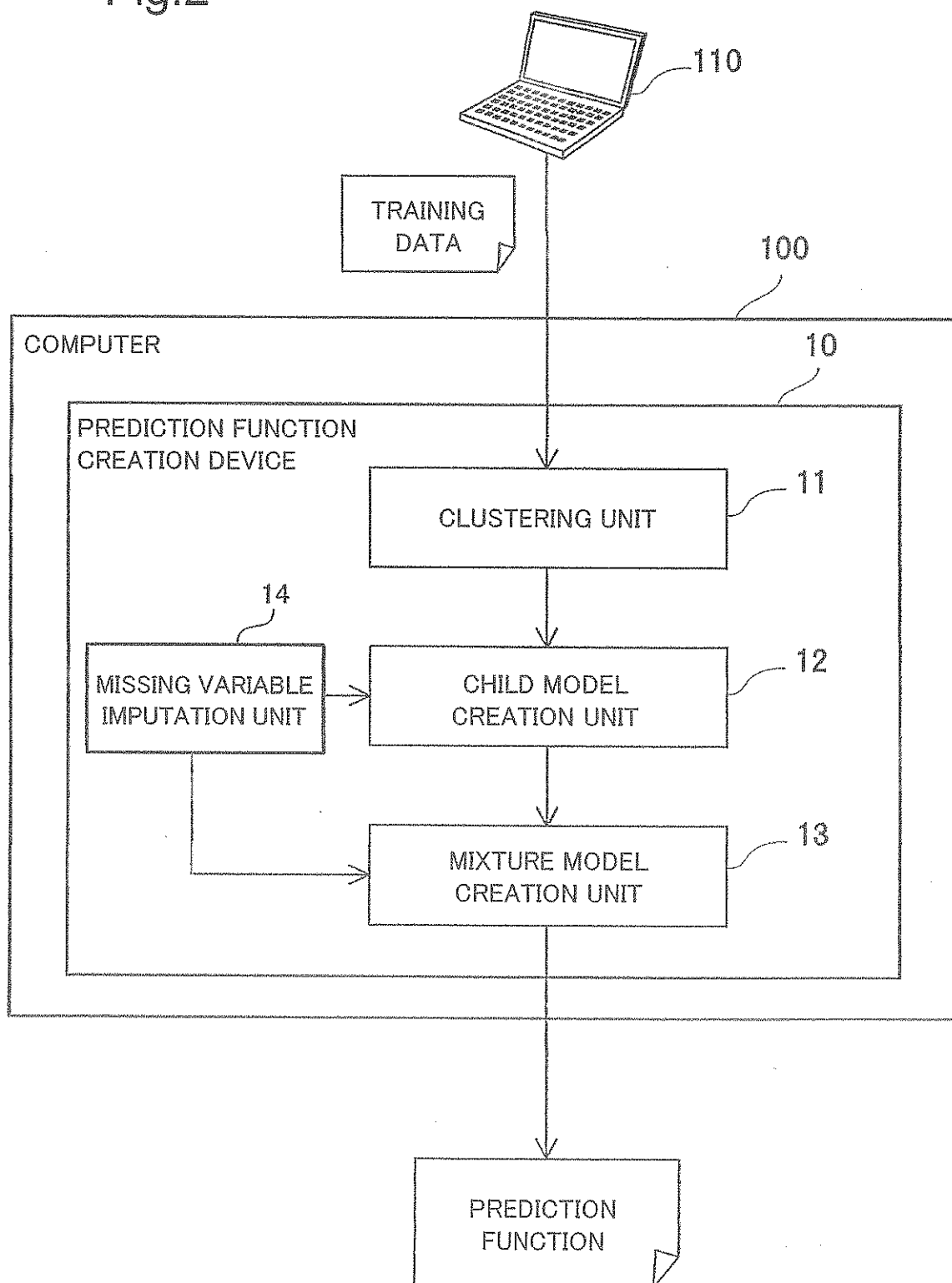


Fig.3

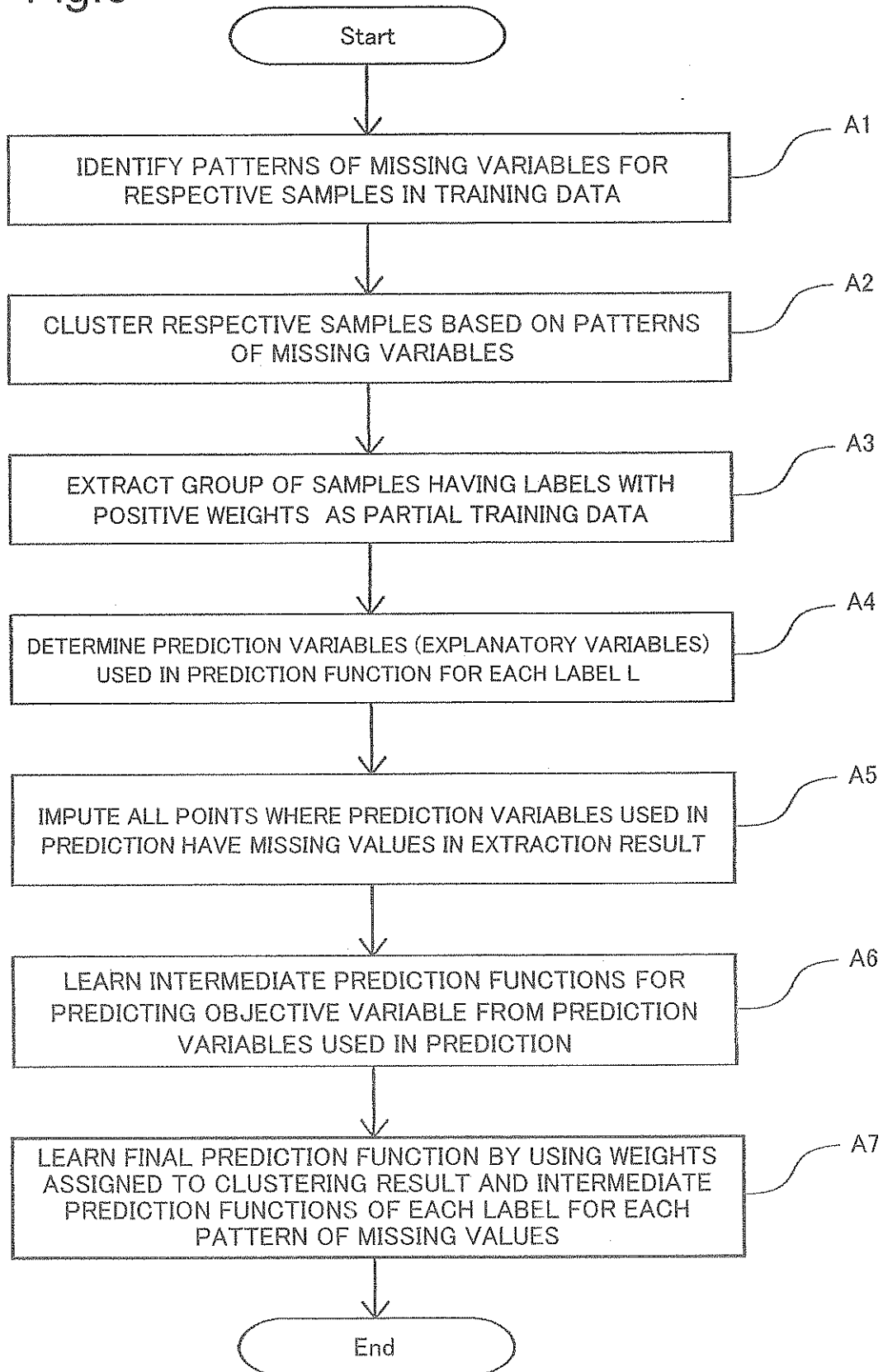


Fig.4

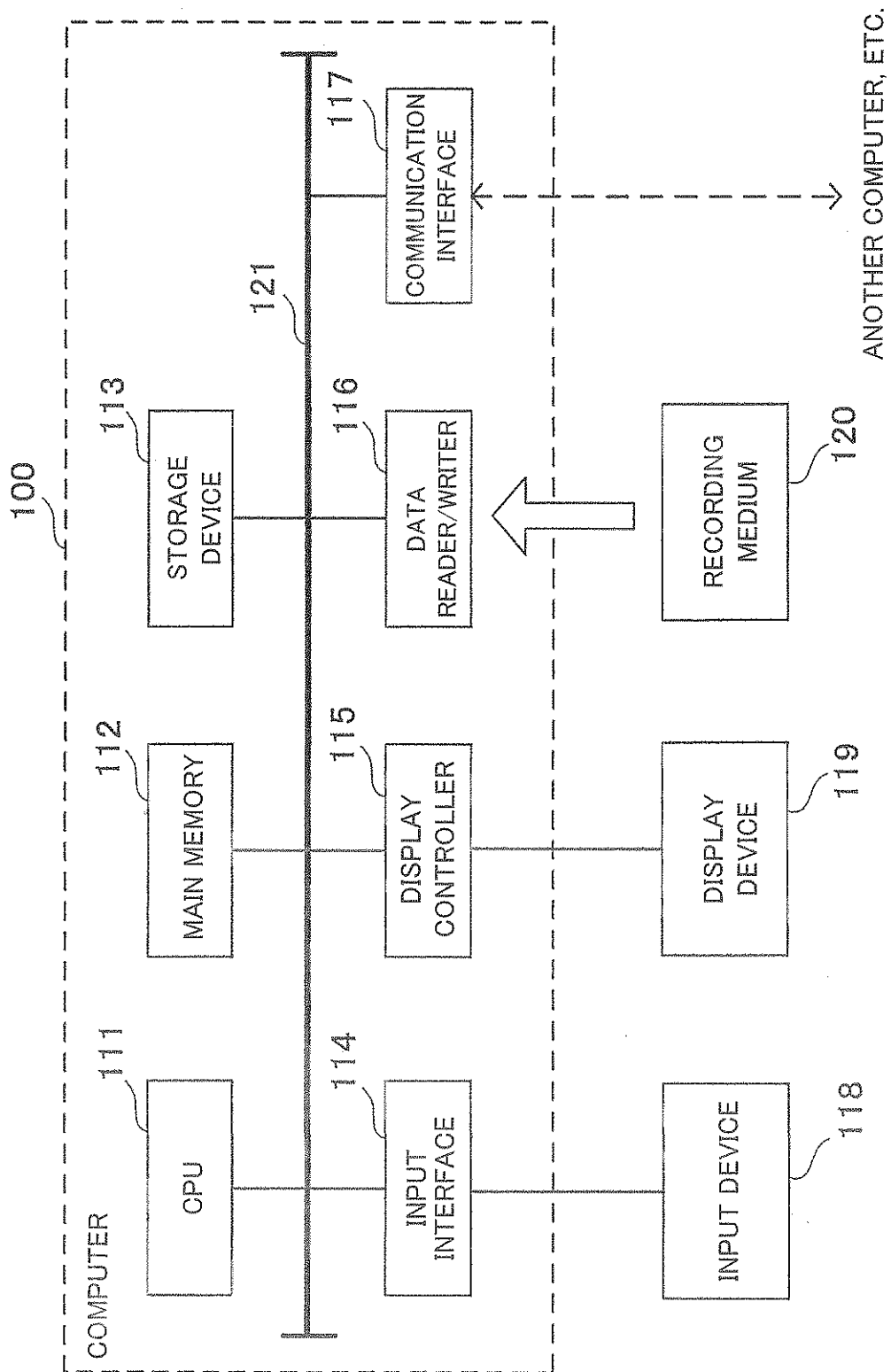


Fig.5

NUMBER	X1	X2	X3	Y
1	*	*	*	*
⋮	*	*	*	*
5	*	*	*	*
6	NA	*	*	*
⋮	NA	*	*	*
10	NA	*	*	*
11	NA	NA	*	*
⋮	NA	NA	*	*
20	NA	NA	*	*
21	*	*	NA	*
⋮	*	*	NA	*
40	*	*	NA	*

Fig.6

SET OF NON-MISSING VARIABLES	MULTIPLIED VALUE
EMPTY SET	0
X1	25
X2	30
X3	20
X1, X2	50
X2, X3	20
X1, X3	10
X1, X2, X3	15

Fig.7

PATTERN OF MISSING VARIABLES	CANDIDATE (X1)	CANDIDATE (X2)	CANDIDATE (X3)	CANDIDATE (X1, X2)	CANDIDATE (X2, X3)
EMPTY SET	2	2	2	1	1
X1	4	1	1	3	0
X1, X2	3	3	0	5	2
X3	2	1	4	0	3
X2	1	4	1	3	3
X2, X3	0	3	3	2	5
X1, X3	3	0	3	2	2

Fig.8

PATTERN OF MISSING VARIABLES	LABEL 1 (X2)	LABEL 2 (X3)	LABEL 3 (X1 , X2)	LABEL 4 (X2 , X3)	LABEL 5 (X1)
EMPTY SET	-	-	0.5	0.5	-
X1	0.21	0.21	-	0.58	-
X1, X2	-	1	-	-	-
X3	0.27	-	0.73	-	-
X2	-	0.5	-	-	0.5
X2, X3	-	-	-	-	1
X1, X3	1	-	-	-	-

Fig.9

LABEL	SAMPLE NUMBER
LABEL 1 (X2)	1-10, 21-40
LABEL 2 (X3)	1-20
LABEL 3 (X1, X2)	1-5, 21-40
LABEL 4 (X2, X3)	1-10
LABEL 5 (X1)	1-5, 21-40

Fig.10

PATTERN OF MISSING VARIABLES	FINAL PREDICTION FUNCTION
EMPTY SET	$0.5 f_3(x_1, x_2) + 0.5 f_4(x_2, x_3)$
X1	$0.21 f_1(x_2) + 0.21 f_2(x_3) + 0.58 f_4(x_2, x_3)$
X1, X2	$f_2(x_3)$
X3	$0.27 f_1(x_2) + 0.73 f_3(x_1, x_2)$
X2	$0.5 f_2(x_3) + 0.5 f_5(x_1)$
X2, X3	$f_5(x_1)$
X1, X3	$f_1(x_2)$

Fig.11

