

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;