

DATA COMPLEMENTING SYSTEM AND DATA COMPLEMENTING METHOD

CROSS-REFERENCE TO RELATED APPLICATION

[0001]

This application claims priority based on Japanese patent application, No. 2020-040800 filed on March 10, 2020, the entire contents of which are incorporated herein by reference.

BACKGROUND

Technical Field

[0002]

The present invention relates to a data complementing system and a data complementing method.

Related Art

[0003]

Japanese Patent No. 6472589 discloses a map data processing apparatus that complements map data of a data missing region in which attribute data is missing, in the map data including the attribute data of features, which is used in various operations such as various planning and business evaluation by business entities such as local governments, retailers and delivery companies. The map data processing apparatus searches a plurality of regions, and acquires similar regions similar to the data missing region.

The map data processing apparatus generates complemented map data in which the missing data is completed in the map data of the data missing region, based on the map data of the similar region. The map data processing apparatus analyzes data based on the map data including the complemented map data, and determines suitability of the analysis result based on the complemented map data. When the determination result is appropriate, the map data processing apparatus outputs the complemented map data.

[0004]

The map data processing apparatus disclosed in Japanese Patent No. 6472589 complements the missing data of the map data including the attribute data of the feature of the data missing region, based on the data of the region similar to the data missing region. Therefore, in a case where there is no region similar to the data missing region, there is a possibility that it is not possible to complement the missing data.

SUMMARY

[0005]

An object of the present invention is to provide a data complementing system and a data complementing method capable of efficiently complementing missing data in data regarding a region.

[0006]

To achieve the above object, an aspect of the present invention provides a data complementing system. The data complementing system includes a storage unit and a complement model generation unit. The storage unit is configured to store region characteristic data that includes values of a plurality of data items regarding a predetermined region, cell-region characteristic data that includes values of a plurality of data items regarding a cell region that is a region obtained by dividing the region into a mesh, information indicating a missing data item that is the data item of missing data being data missed in the cell-region characteristic data, external region characteristic data that includes values of a plurality of data items regarding an external region that is different from the region, and an external cell-region characteristic data that includes values of a plurality of data items regarding an external cell region obtained by dividing the external region into a mesh. The complement model generation unit is configured to generate a complement model for generating complement data that is for complementing the missing data, based on the external region characteristic data and the external cell-region characteristic data.

[0007]

In addition, the details of one or more implementations of the subject matter described in the specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.

[0008]

According to the present disclosure, it is possible to complement data that is missing in data regarding a region, with high efficiency.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009]

Fig. 1 is a diagram illustrating a schematic configuration of a data complementing system to be described as an embodiment;

Fig. 2A is a diagram illustrating an example of a region;

Fig. 2B is a diagram illustrating an example of a cell region;

Fig. 3A is a diagram illustrating an example of region characteristic data;

Fig. 3B is a diagram illustrating an example of cell-region characteristic data;

Fig. 4 illustrates an example of a hardware

configuration of an information processing apparatus used in a configuration of the data complementing system;

Fig. 5 is a flowchart illustrating an example of a complement data generation process;

Fig. 6 illustrates an example of a user information setting screen;

Fig. 7 illustrates an example of an explanatory variable setting screen;

Fig. 8 illustrates an example of a region-unit data check screen;

Fig. 9 is a schematic diagram illustrating an example of model generation data;

Fig. 10 is a flowchart illustrating an example of a complement model generation process;

Fig. 11A illustrates an example of external cell-region characteristic data;

Fig. 11B illustrates an example of external region characteristic data;

Fig. 12 is a flowchart illustrating an example of a cell-region characteristic data multiple regression analysis process;

Fig. 13 is a flowchart illustrating an example of a region characteristic data multiple regression analysis process;

Fig. 14A is a diagram illustrating an example of cell-

region characteristic partial regression information;

Fig. 14B is a diagram illustrating an example of region-characteristic partial regression information;

Fig. 15 is a diagram illustrating an example of a use complement model selection screen;

Fig. 16 is a diagram illustrating an example of a complement information confirmation screen;

Fig. 17 is a diagram illustrating an example of complement model information; and

Fig. 18 is a diagram illustrating an example of complement data information.

DESCRIPTION OF EMBODIMENTS

[0010]

Hereinafter, an embodiment will be described with reference to the drawings. In the following description, the same or similar components may be denoted by the same reference signs, and repetitive description may be omitted. In the following description, the letter "s" in front of the reference sign means a processing step.

[0011]

Fig. 1 illustrates a schematic configuration of an information processing system (referred to as a "data complementing system 1" below) which will be described as an embodiment. A user of the data complementing system 1 is,

for example, the local government or a service provider.

[0012]

When planning and implementing services for regions such as cities, wards, and blocks, local governments and service providers collect region characteristic data that includes various types of information (various data items and values of the data items) regarding a region as a target of a service, and perform a work such as data analysis using the collected region characteristic data. Examples of the data items include the area, the number of households, the population, the age distribution, and the number of registered automobiles in a region.

[0013]

When performing the work using the region characteristic data, the data complementing system 1 generates complement data for complementing missing data which is data missed in the region characteristic data.

[0014]

For the purpose of providing high-quality services, local governments and service providers divide a region into a plurality of mesh-like (for example, 500 m × 500 m) regions (referred to as a "cell region") below, and handle cell-region characteristic data being various types of information (various data items and values of the data items) regarding the cell region, along with region characteristic data.

Examples of the data items include the area, the number of households, the population, and the age distribution in a cell region.

[0015]

Figs. 2A and 2B illustrate the concept of a region 300 and a cell region 301. Fig. 2A is a diagram illustrating the entirety of the region 300 such as a city. Fig. 2B is an enlarged view of a rectangular region 302 illustrated in Fig. 2A.

[0016]

Fig. 3A is a diagram illustrating an example of region characteristic data 110. As illustrated in Fig. 3A, the region characteristic data 110 is data in which, for a region (identifier of the region), the area of the region and values of a plurality of data items (population, age population under 15, age population over 65, number of owned vehicles, and the like) are associated with each other.

[0017]

Fig. 3B illustrates an example of cell-region characteristic data 120. As illustrated in Fig. 3B, the cell-region characteristic data 120 is a data in which, for a cell region (identifier of the cell region), a cell size representing the area of the cell region and values of a plurality of data items (population, age population under 15, age population over 65, and the like) are associated with

each other.

[0018]

Some data items (for example, number of owned vehicles in the region characteristic data 110 illustrated in Fig. 3A) are provided in the region characteristic data 110, but are not provided in the cell-region characteristic data 120. The value of a missing data item, which is such a data item not provided in the cell-region characteristic data 120, is the above-described missing data. The data complementing system 1 generates complement data for such missing data.

[0019]

The data complementing system 1 generates the complement data based on external region characteristic data or external cell-region characteristic data. The external region characteristic data is data including various types of information (various data items and values of the data items) regarding an external region that is a region other than a region as a target of a service. The external cell-region characteristic data refers to various types of information (various data items and values of the data items) regarding an external cell region that is a region obtained by dividing the external region into a mesh (for example, 500 m × 500 m).

[0020]

As illustrated in Fig. 1, the data complementing

system 1 includes a user terminal 20 and a data complementing apparatus 100. Both the user terminal 20 and the data complementing apparatus 100 are configured using an information processing apparatus (computer), and are connected to each other to be able to communicate with each other via a wired or wireless communication network 30. The communication network 30 is, for example, a local area network (LAN), a wide area network (WAN), the Internet, a dedicated line, and various data communication networks.

[0021]

The user terminal 20 is managed by, for example, a local government or a service provider who plans and implements a service. The user terminal is operated by a person (referred to as a "user" below) who performs the above-described work. The data complementing apparatus 100 provides the user terminal 20 with various services related to complementing missing data.

[0022]

Fig. 4 illustrates an example of a hardware configuration of an information processing apparatus 10 constituting the user terminal 20 or the data complementing apparatus 100. As illustrated in Fig. 4, the information processing apparatus 10 includes a processor 11, a main storage device 12, a communication interface 13, and an auxiliary storage device 14.

[0023]

The processor 11 is configured using, for example, a central processing unit (CPU), a micro processing unit (MPU), a graphics processing unit (GPU), an artificial intelligence (AI) chip, a field programmable gate array (FPGA), a system on chip (SoC), and an application specific integrated circuit (ASIC).

[0024]

The main storage device 12 stores programs and data, and includes, for example, a read only memory (ROM), a random access memory (RAM), and a non-volatile memory (NVRAM).

[0025]

The communication interface 13 communicates with another information processing apparatus via the communication network 30, and includes a wireless or wired communication module (wireless communication module, communication network adapter, USB module, and the like).

[0026]

The auxiliary storage device 14 stores programs and data, and includes, for example, a solid state drive (SSD), a hard disk drive, an optical storage medium (compact disc (CD), digital versatile disc (DVD), and the like), an IC card, and an SC card. The auxiliary storage device 14 stores programs and data for realizing the functions of the data complementing apparatus 100. The auxiliary storage device

14 may write and read programs and data via a reading device of a recording medium or the communication interface 13. Programs and data stored in the auxiliary storage device 14 are read out to the main storage device 12 at any time.

[0027]

An input device 15 is a user interface that receives a user input and a data input from the outside of the information processing apparatus. For example, the input device includes a keyboard, a mouse, a touch panel, a card reader, and a voice input device (for example, a microphone).

[0028]

An output device 16 is a user interface that outputs various types of information to the user, and includes a display device (liquid crystal display, organic EL panel, and the like) that displays various types of information, an audio output device (for example, speaker) that outputs various types of information by audio, a printer that performs printing on a paper medium, and the like.

[0029]

The information processing apparatus 10 includes, for example, a personal computer (desktop type or notebook type), a smartphone, a tablet, and a general-purpose machine. The information processing apparatus 10 may be realized using virtual information processing resources such as a cloud server provided by a cloud system, for example. The

information processing apparatus 10 may be configured by a plurality of information processing apparatuses distributed on a communication network. For example, software for realizing an operating system, a file system, a database management system (DBMS) (relational database, NoSQL, and the like), a key-value store (KVS), or the like may be installed on the information processing apparatus 10.

[0030]

Various functions of the user terminal 20 or the data complementing apparatus 100 are realized in a manner that the processor 11 of the information processing apparatus 10 constituting the user terminal or the data complementing apparatus reads and executes one or more programs stored in the main storage device 12. Alternatively, the various functions are realized by the hardware of the information processing apparatus 10 constituting the user terminal or the data complementing apparatus. The program may be stored in the auxiliary storage device 14 in advance. If necessary, the program may be stored in the auxiliary storage device 14 from a non-temporary storage device of another apparatus via the communication network 30, or from a non-temporary storage medium.

[0031]

As illustrated in Fig. 1, the user terminal 20 includes the functions of a Web browser 21, a communication unit 22,

a user interface unit 23, and a storage unit 24. The Web browser 21 displays information transmitted from the data complementing apparatus 100, transmits information received from the user to the data complementing apparatus 100, and the like. The communication unit 22 realizes a communication with the data complementing apparatus 100 via the communication network 30. The user interface unit 23 realizes interactive processing with the user via the input device 15 and the output device (display device) 16.

[0032]

The data complementing apparatus 100 includes the functions of a storage unit 200, a communication unit 101, a screen data generation unit 102, a user setting-information acquisition unit 103, a data acquisition unit 104, a complement model generation unit 105, a data complementing unit 106, a complement model selection unit 107, and a complement data correction unit 108.

[0033]

Among the above functions, the storage unit 200 stores data of each of user setting information 210, region characteristic data 110, cell-region characteristic data 120, model generation data 130, external cell-region characteristic data 220, external region characteristic data 230, cell-region-characteristic partial regression information 240, region-characteristic partial regression

information 250, complement model information 260, and complement data information 270. The storage unit 200 stores each type of data, for example, as a database table provided by the DBMS, a file provided by a file system, and KVS data. Details of the data will be described later.

[0034]

The communication unit 101 communicates with the user terminal 20 via the communication network 30.

[0035]

The screen data generation unit (output unit) 102 functions as a Web server. The screen data generation unit transmits, for example, data (for example, image data or script for realizing the user interface) described in a format such as the hypertext markup language (HTML), the extensible markup language (XML), and the JavaScript (registered trademark) object notation (JSON), to the Web browser 21 of the user terminal 20 via the communication unit 101.

[0036]

The user setting-information acquisition unit (receiving unit) 103 acquires information (referred to as "user setting information" below) transmitted from the user terminal 20. The user setting-information acquisition unit 103 may be realized as a function of the above-described Web server. The storage unit 200 stores the user setting

information acquired by the user setting-information acquisition unit 103 as the user setting information 210. The user setting-information acquisition unit 103 acquires region characteristic data and cell-region characteristic data, which are transmitted from the user terminal 20. The storage unit 200 stores the pieces of data acquired by the user setting-information acquisition unit 103, as the region characteristic data 110 and the cell-region characteristic data 120, respectively.

[0037]

The data acquisition unit 104 acquires external region characteristic data and external cell-region characteristic data from an accessible information source (database, Web server, and the like which are accessible via the communication network 30) via the communication network 30. The storage unit 200 stores the pieces of data acquired by the data acquisition unit 104, as the external region characteristic data 230 and the external cell-region characteristic data 220, respectively.

[0038]

The complement model generation unit 105, the data complementing unit 106, the complement model selection unit 107, and the complement data correction unit 108 generate a complement model based on the data acquired by the user setting-information acquisition unit 103 and the data

acquisition unit 104, and generate complement data based on the generated complement model. The complement model is a model for generating information which is for complementing (predicting) the above-described missing data. Details of the functions will be described later.

[0039]

Fig. 5 is a flowchart illustrating a process (referred to as a "complement data generation process s5" below) of generating complement data by the data complementing apparatus 100. The complement data generation process s5 will be described with reference to Fig. 5.

[0040]

Firstly, the user setting-information acquisition unit 103 in the data complementing apparatus 100 transmits a screen (referred to as a "user information setting screen 40" below) for setting the above-described missing data items and the like, to the user terminal 20. The user terminal 20 receives and displays the user information setting screen 40. The user inputs a missing data item (also referred to as a "complement data item" below) desired to be complemented on the user information setting screen 40 (s11). The details of the user information setting screen 40 will be described later.

[0041]

Then, the user sets data (cell-region characteristic data, and region characteristic data) possessed by the user, which can be used for generating the complement data, on the user information setting screen 40 (s12).

[0042]

The data complementing apparatus 100 transmits a screen (referred to as an "explanatory variable setting screen 50" below) for setting a data item group (referred to as an "explanatory variable pattern" below) of the region characteristic data used as an explanatory variable when generating a complement model. The user terminal 20 receives and displays the explanatory variable setting screen 50. The user sets one or more explanatory variable patterns by performing a predetermined input operation such as selection of one or more data items as explanatory variables on the explanatory variable setting screen 50 one or more times (s13). The details of the explanatory variable setting screen 50 will be described later.

[0043]

Then, the user terminal 20 transmits the information (user setting information) set in s11 to s13, the cell-region characteristic data, and the region characteristic data to the data complementing apparatus 100 via the communication network 30 (s14). In the transmission, the user terminal 20 displays a screen (referred to as a "region-unit data check

screen 60" below) for checking whether or not the user terminal stores region-unit data (whether or not the user terminal is capable of providing the region-unit data), so as to check whether or not the user has the region-unit data. The region-unit data refers to region characteristic data including the value of the missing data item. The user terminal 20 transmits the result of the above check to the data complementing apparatus 100 along with the above information.

[0044]

The user setting-information acquisition unit 103 in the data complementing apparatus 100 receives the user setting information, the cell-region characteristic data, the region characteristic data, and the result of the above check, which are transmitted from the user terminal 20 (s21). The data complementing apparatus 100 determines the result of the above check (s22). When the user has the region-unit data (s22: YES), the data complementing apparatus performs processing from s24. When the user does not have the region-unit data (s22: NO), the data complementing apparatus 100 performs processing from s23.

[0045]

In s23, the data complementing apparatus 100 generates region-unit complement data for the received region characteristic data. The region-unit complement data is

region characteristic data in which the value of the missing data item is complemented in units of regions. The data complementing apparatus 100 generates the region-unit complement data by a method of, for example, replacing the missing data item with data having a meaning similar to that of the missing data item among items of the region characteristic data 110 of another region, which are possessed by the user. The process of s23 may not be necessarily executed (that is, the region-unit complement data may not be generated). In this case, the process proceeds to the next process of s24. As a case where the process of s23 is not performed, for example, the data complementing apparatus 100 may determine that it is not possible to generate region-unit complement data with the required accuracy, based on information of region characteristic data possessed by the user, which indicates, for example, that external region characteristic data having similar values of the data items other than the missing data item (complement data item) is not provided.

[0046]

The processes of s24 to s27 are processes (loop processes) that are repeatedly executed while sequentially selecting explanatory variable patterns. In s24, the data complementing apparatus 100 selects one explanatory variable pattern that has not yet been targeted by the loop processes

s25 to s26 from one or more explanatory variable patterns set by the user in s13.

[0047]

The complement model generation unit 105 in the data complementing apparatus 100 performs a process (referred to as a "complement model generation process s25" below) of generating the complement model, that is, generating the complement model by using the items of the cell-region characteristic data 120 and the items of the region characteristic data 110, which correspond to the explanatory variables belonging to the explanatory variable pattern selected in s24. The information regarding the complement model generated by the complement model generation process s25 is stored in the storage unit 200 as the complement model information 260. The details of the complement model generation process s25 will be described later.

[0048]

The data complementing unit 106 in the data complementing apparatus 100 substitutes the values of the region characteristic data 110 and the cell-region characteristic data 120 into the complement model generated by the complement model generation process s25 to obtain the value of the missing data item. Then, the data complementing unit generates a cell-unit complement data which is complement data in units of cells, by using the obtained

values (s26). The generated cell-unit complement data is stored in the storage unit 200 as the complement data information 270.

[0049]

The data complementing unit 106 generates region conversion complement data being data obtained by converting the value of the cell-unit complement data of all the cell regions of the region into the value in units of the regions by performing processing such as obtaining the total or the average. The data complementing unit 106 obtains the deviation between the region conversion complement data and the region-unit data (region-unit complement data when the region-unit complement data is generated in s23). The data complementing unit 106 obtains the above deviation from, for example, the following equation.

[0050]

Deviation = (value of region conversion complement data - value of region-unit data (value of region-unit complement data))/value of region-unit data (value of region-unit complement data) ... Equation 1

[0051]

When the user does not have the region-unit data and does not calculate the region-unit complement data, the deviation is "no value".

[0052]

The data complementing apparatus 100 determines whether or not the loop processes s25 to s26 have been completed for all the explanatory variable patterns set in s13. When the loop processing s25 to s26 have not been completed for all explanatory variable patterns (s27: NO), processing returns to s24. If loop processing s25 to s26 is completed for all explanatory variable patterns (s27: YES), processing proceeds to s28.

[0053]

In s28, the data complementing apparatus 100 transmits the information regarding the complement model generated in s25 to s26 and the information regarding the complement data, to the user terminal 20 via the communication network 30.

[0054]

The data complementing apparatus 100 transmits a screen (referred to as a "use complement model selection screen 70" below) for causing the user to select the use complement model, to the user terminal 20. The user terminal 20 receives and displays the use complement model selection screen 70, and receives, from the user, a designation of a complement model (referred to as a "use complement model" below) to be used by the user (s15). When the user has the region-unit data, the user selects whether or not to correct the "cell-unit complement data" complemented using the use

complement model, on the use complement model selection screen 70.

[0055]

Then, the user terminal 20 transmits the information of the use complement model selected in s15 and information (referred to as "correction necessity information" below) indicating whether or not to correct the "cell-unit complement data" complemented using the selected use complement model, to the data complementing apparatus 100 via the communication network 30 (s16).

[0056]

When the data complementing apparatus 100 receives the information of the use complement model and the correction necessity information, which are transmitted from the user terminal 20, the data complementing apparatus generates "confirmed cell-unit complement data" being data in which contents of the cell-unit complement data are confirmed, based on the received information (s29). More specifically, the data complementing apparatus 100 firstly checks the correction necessity information. When the correction necessity information indicates at least any of a case of "not corrected", a case where the user does not have the region-unit data, and a case where the use complement model is selected based on the significance, the data complementing apparatus 100 sets the "cell-unit complement data" itself

generated using the use complement model, as the confirmed cell-unit complement data. When the correction necessity information indicates "correct", the complement data correction unit 108 in the data complementing apparatus 100 corrects the cell-unit complement data generated using the use complement model selected by the user in s15, so as to generate the confirmed cell-unit complement data. Specifically, the complement data correction unit 108 generates the confirmed cell-unit complement data by, for example, calculating a coefficient from the deviation obtained in s26 and multiplying the calculated coefficient by the cell-unit complement data.

[0057]

The data complementing apparatus 100 transmits the confirmed cell-unit complement data generated in s29, to the user terminal 20 via the communication network 30 (s30).

[0058]

When the user terminal 20 receives the confirmed cell-unit complement data transmitted from the data complementing apparatus 100, the user terminal displays a screen (referred to as a "complement information confirmation screen 80" below) on which the received confirmed cell-unit complement data, the missing data item (complement data item) registered by the user in s11, the information regarding the use

complement model selected by the user in s15, and the like are described (s17).

[0059]

<Detailed Description>

The complement data generation process s5 in Fig. 5 will be described in more detail.

[0060]

Fig. 6 illustrates an example of the user information setting screen 40 displayed by the user terminal 20 in the processes of s11 to s13 of Fig. 5. As illustrated in Fig. 6, the user information setting screen 40 includes a complement data item input field 41, a cell-region characteristic data input field 42, a cell size input field 43, a region characteristic data input field 44, and a complement model significance level input field 45.

[0061]

In the complement data item input field 41, information (name of the missing data item in this example) for identifying the missing data item (complement data item) of the missing data desired to be complemented by the user among the missing data items of the cell-region characteristic data is input.

[0062]

In the cell-region characteristic data input field 42, information (for example, information indicating the file

name and location of the cell-region characteristic data) for identifying the cell-region characteristic data possessed by the user is input.

[0063]

In the cell size input field 43, the size (length of one side of a cell in the mesh in this example) of the cell region is input.

[0064]

In the region characteristic data input field 44, information that specifies the region characteristic data possessed by the user (for example, information indicating the file name and location of the cell-region characteristic data) is input.

[0065]

In the complement model significance level input field 45, a significance level that serves as a criterion for determining the significance of the complement model generated by the complement model generation unit 105 in the data complementing apparatus 100 is input.

[0066]

When the user operates a registration button 46, the user terminal 20 starts reception of the explanatory variable pattern (s13 in Fig. 5).

[0067]

Fig. 7 illustrates an example of the explanatory variable setting screen 50 displayed by the user terminal 20 in s13 of Fig. 5. As illustrated in Fig. 7, the explanatory variable setting screen 50 includes an explanatory variable pattern number display field 51, an explanatory variable selection field 52, a "next" button 54, and a pattern registration end button 55.

[0068]

An identifier (referred to as a "pattern number" below) of the currently-displayed explanatory variable pattern is set in the explanatory variable pattern number display field 51. The pattern number displayed in the explanatory variable pattern number display field 51 is updated in the order of 1, 2, 3,... each time the "next" button 54 is instructed.

[0069]

In the explanatory variable selection field 52, data items 53 included in the cell-region characteristic data input on the user information setting screen 40 are listed in a state in which the data items can be selected by check boxes. The data item 53 selected by the user in the explanatory variable selection field 52 is used as an explanatory variable in the multiple regression analysis described later.

[0070]

When the user operates the pattern registration end button 55, the user terminal 20 displays the region-unit data check screen 60 (s14 in Fig. 5). The region-unit data check screen 60 is a screen for causing the user to check whether or not the missing data item (complement data item) input on the user information setting screen 40 is included in the region characteristic data, and for receiving an instruction to transmit the user setting information.

[0071]

Fig. 8 illustrates an example of the region-unit data check screen 60 displayed by the user terminal 20 in the process of s14 of Fig. 5. As illustrated in Fig. 8, the region-unit data check screen 60 includes a list of data items 61 included in the region characteristic data, and a black circle mark 62 indicating the complement data items designated as being possessed by the user in the list 61, and a checkmark 63 indicating that the user has a complement data item in units of regions (indicating that the complement data item is included in the region characteristic data). When the user operates a confirmation button 64, the user terminal 20 transmits the user setting information set or acquired by the above processing to the data complementing apparatus 100.

[0072]

Fig. 9 illustrates an example of information (referred to as "model generation data 130" below) transmitted by the user terminal 20 to the data complementing apparatus 100 as the above user setting information. As illustrated in Fig. 9, the model generation data 130 includes the contents of the explanatory variable pattern set by the user on the explanatory variable setting screen 50 (contents in which the pattern number is associated with the data item set as the explanatory variable with the pattern number). In the illustrated model generation data 130, "o" is set for the data item selected by the user, and "x" is set for the data item not selected by the user. Each explanatory variable pattern is uniformly associated with the complement data item "number of owned vehicles" and the significance level "5%".

[0073]

Fig. 10 is a flowchart illustrating the details of the complement model generation process s25 in Fig. 5. The complement model generation process s25 will be described below with reference to Fig. 10.

[0074]

The data acquisition unit 104 firstly acquires the external region characteristic data 230 including the data item and the complement data item designated by one selected explanatory variable pattern and external cell-region characteristic data 220, from an accessible information

source (s31). The storage unit 200 stores the external region characteristic data 230 and the external cell-region characteristic data 220, which are acquired by the data acquisition unit 104.

[0075]

Fig. 11A illustrates an example of the external cell-region characteristic data 220. The external cell-region characteristic data 220 is data in which various data items are associated with the external cell region (identifier of the external cell region). The external cell-region characteristic data 220 includes a data item corresponding to the complement data item designated by the user among the missing data items in the cell-region characteristic data. The illustrated external cell-region characteristic data 220 includes the pattern number of the explanatory variable pattern selected by the complement model generation unit 105.

[0076]

Fig. 11B illustrates an example of the external region characteristic data 230. The external region characteristic data 230 is data in which data items are associated with an external region (identifier of the external region). The external region characteristic data 230 includes data items corresponding to complement data items.

[0077]

Returning to Fig. 10, the complement model generation unit 105 performs multiple regression analysis (referred to as a "cell-region characteristic data multiple regression analysis process s32" below) on the external cell-region characteristic data 220 of the external cell region belonging to each external region extracted by the data acquisition unit 104.

[0078]

Fig. 12 is a flowchart illustrating an example of the cell-region characteristic data multiple regression analysis process s32. Firstly, the complement model generation unit 105 selects one external region (s51), and acquires the external cell-region characteristic data 220 of an external cell region belonging to the selected external region, as the target of multiple regression analysis (s52).

[0079]

The complement model generation unit 105 performs multiple regression analysis (referred to as "external cell-region characteristic multiple regression analysis" below) on the acquired external cell-region characteristic data 220 (s53). A regression equation (first regression equation) by the external cell-region characteristic multiple regression analysis is, for example, as follows.

$$y_A = b_{0A} + x_{1A}b_{1A} + x_{2A}b_{2A} + \dots + x_{nA}b_{nA} \dots \text{Equation 2}$$

[0080]

In the above equation, the subscripts y , b , and x indicate the identifiers of the external regions. In the above equation, the multiple regression analysis is performed for an external region A. The complement model generation unit 105 uses the data item (for example, number of owned vehicles) corresponding to the missing data item in the external cell-region characteristic data 220 with an objective variable as y_A , and uses the explanatory variables in the explanatory variable pattern with the explanatory variables as x_{1A} , x_{2A} , and the like. For example, when the explanatory variable pattern in which the pattern number in the model generation data 130 illustrated in Fig. 9 is "p02" is provided, the explanatory variables x_{1A} , x_{2A} , x_{3A} , ... to be used are the population, the age population under 15, the age population over 65, and the like.

[0081]

The complement model generation unit 105 uses the objective variable and the explanatory variable in the external cell-region characteristic data 220 in the external region A to obtain b_0 (intercept) in the above regression equation and b_i ($i=1$ to n) which is a partial regression coefficient of each explanatory variable. In the following description, b_0 (intercept) and the partial regression coefficient b_i ($i=1$ to n) are collectively referred to as a

cell-region characteristic partial regression coefficient b_i ($i = 0$ to n).

[0082]

Returning to Fig. 10, the complement model generation unit 105 tests the significance of the above regression equation by an F test or the like (referred to as a "cell analysis significance test").

[0083]

The complement model generation unit 105 tests the significance of all the cell-region characteristic partial regression coefficients b_i in order from $i = 0$ to n one by one. The complement model generation unit 105 firstly sets i to 0 (s33). Then, the complement model generation unit 105 obtains a difference between all the cell-region characteristic partial regression coefficient b_i from the cell-region characteristic partial regression information by the following equation (s34), for all combinations of two external regions selected from all the external regions acquired by the data acquisition unit 104.

$$\Delta b_{0AB} = b_{0B} - b_{0A} \dots \text{Equation 3}$$

[0084]

The above equation corresponds to a case where an external region A and an external region B are selected as the two external regions. In the above equation, a

difference Δb_{0AB} is obtained by using the cell-region characteristic partial regression coefficient b_0 (intercept) as the calculation target of the difference. The complement model generation unit 105 obtains the difference Δb_i between all the cell-region characteristic partial regression coefficients b_i for all combinations of the two external regions by the above equation.

[0085]

The complement model generation unit 105 tests a hypothesis of "the difference of the cell-region characteristic partial regression coefficient is 0" by using the difference Δb_i of the cell-region characteristic partial regression coefficients b_i for all combinations of the two external regions selected from all the extracted external regions (s35). When the complement model generation unit 105 determines that the above hypothesis is statistically "significant" (s35: YES), the complement model generation unit obtains the mean value of the cell-region characteristic partial regression coefficient b_i (s36). The complement model generation unit 105 uses the value of the significance level included in the model generation data 130 in this test. The complement model generation unit 105 determines the significance based on the test result. The test result of the significance is reflected in the numerical value of the significance ratio of "cell analysis" on the use complement

model selection screen 70, which will be described later along with Fig. 15.

[0086]

When the complement model generation unit 105 determines that the above hypothesis is statistically "not significant" (s35: NO), the complement model generation unit performs multiple regression analysis (referred to as a "region characteristic data multiple regression analysis process s37" below) based on the difference Δb_i of the cell-region characteristic partial regression coefficient b_i and the information included in the external region characteristic data 230.

[0087]

Fig. 13 is a flowchart illustrating an example of the region characteristic data multiple regression analysis process s37. Firstly, the complement model generation unit 105 obtains the difference of the external region characteristic data 230 from the following equation, for all combinations for two external regions selected from the firstly-extracted external regions (s61).

$$\Delta D_{1AB} = D_{1B} - D_{1A} \dots \text{Equation 4}$$

[0088]

The above equation is an equation for calculating the difference ΔD_{1AB} between the values D_{1A} and D_{1B} of the same

data item D_1 in external region characteristic data of two regions being the external region A and the external region B. For example, in a case of the external region characteristic data 230 illustrated in Fig. 11B, for example, when D_1 indicates the area, D_{1A} and D_{1B} are "32.1" and "14.2", respectively, and $\Delta D_{1AB} = D_{1B} - D_{1A} = 17.9$.

[0089]

The complement model generation unit 105 performs multiple regression analysis, for example by using the difference Δb_i of the cell-region characteristic partial regression coefficient obtained by Equation 3 and the difference ΔD_{1AB} of the external region characteristic data obtained by Equation 4, for the same data item (s62). At this time, the complement model generation unit 105 performs multiple regression analysis by the following regression equation (second regression equation) for all combinations for two external regions selected from the extracted external regions.

$$\Delta b_i = g_1 \Delta D_1 + g_2 \Delta D_2 + \dots + g_m \Delta D_m \dots \text{Equation 5}$$

[0090]

In a case of the external region characteristic data 230 illustrated in Fig. 11B, in Equation 5, for example, the difference Δb_i of the cell-region characteristic partial regression coefficient is used as the objective variable,

and, for example, the difference in area, the difference in the number of complex commercial facilities, and the like are used as the explanatory variable ΔD_j ($j = 1$ to m).

[0091]

The complement model generation unit 105 performs a process (referred to as a "region analysis significance test" below) of testing the significance of the regression equation of Equation 5 by the F test and the like to which the significance level of the model generation data 130 is applied, similar to the cell-region characteristic data multiple regression analysis process s32 illustrated in Fig. 12 (s63). The significance of the regression equation represented by the above equation is reflected in the numerical value of the significance ratio of the "region analysis" included in the use complement model selection screen 70, which will be described later along with Fig. 15.

[0092]

Returning to Fig. 10, when the complement model generation unit 105 executes the above-described cell analysis significance test and region analysis significance test for one cell-region characteristic partial regression coefficient, i is set to $(i+1)$ (s38).

[0093]

The complement model generation unit 105 repeats the loop processes of s34 to s38 until the cell analysis

significance test and the region characteristic data multiple regression analysis process s37 are executed for all the cell-region characteristic partial regression coefficients b_i (s39: NO).

[0094]

When the complement model generation unit 105 executes the cell analysis significance test and the region characteristic data multiple regression analysis process s37 for all the cell-region characteristic partial regression coefficients b_i (s39: YES), then the complement model generation unit 105 generates the cell-region characteristic partial regression information 240 and region-characteristic partial regression information 250 (s40).

[0095]

Fig. 14A illustrates an example of the cell-region characteristic partial regression information 240. As illustrated in Fig. 14A, the illustrated cell-region characteristic partial regression information 240 indicates a data table in which the cell-region characteristic partial regression coefficient b_i in the regression equation of Equation 1 and the determination result of the cell analysis significance test are associated with each other for each of a plurality of external regions. The determination result is set to "o" when the result is "significant" and is set to "x" when the result is "not significant". The significance

ratio of the cell analysis displayed in a significance ratio display field 76 of the use complement model selection screen 70, which will be described later along with Fig. 15, is a ratio of the external region determined to be significant by the cell analysis significance test, to all the external regions as the target of the cell-region characteristic multiple regression analysis.

[0096]

Fig. 14B illustrates an example of the region-characteristic partial regression information 250. As illustrated in Fig. 14B, the illustrated region-characteristic partial regression information 250 indicates a data table in which the partial regression coefficient (region characteristic partial regression coefficient g_i) in the regression equation of Equation 4 and the determination result of the region analysis significance test are associated with each other for each of the cell-region characteristic partial regression coefficient b_i in the regression equation of Equation 1. The significance ratio of the region analysis displayed in a significance ratio display field 76 of the use complement model selection screen 70, which will be described later along with Fig. 15, is a ratio of the cell-region characteristic partial regression coefficient determined to be significant by the region characteristic multiple regression analysis, to all the cell-

region characteristic partial regression coefficients as the target of the region characteristic multiple regression analysis.

[0097]

Returning to Fig. 10, the complement model generation unit 105 generates a complement model in which the complement data item is expressed using the data item included in one explanatory variable pattern, the cell-region characteristic partial regression coefficient, the region characteristic partial regression coefficient, and the external region characteristic information (s41). Equations representing the complement model are shown below.

$$y = b_{0A} + x_1(b_{1A} + \Delta b_1) + x_2(b_{2A} + \Delta b_2) + \dots + x_n(b_{nA} + \Delta b_n) \dots \text{Equation 6}$$

[0098]

Δb_j ($j = 1$ to n) in Equation 5 is represented by Equation 7.

$$\Delta b_j = g_{1j}(D_1 - D_{1A}) + g_{2j}(D_2 - D_{2A}) + \dots \dots \text{Equation 7}$$

[0099]

In the above equation, b_{0A} , b_{1A} , b_{2A} ,... indicate, for example, cell-region characteristic partial regression coefficients of the region A. D_{1A} , D_{2A} ,... indicate, for example, values of the data items designated by the

explanatory variable patterns in the external region characteristic data of the region A.

[0100]

In s26 of Fig. 5, the data complementing unit 106 substitutes the value of the data item in the cell-region characteristic data 120 designated as the explanatory variable into x_i of the equation of the complement model represented by Equation 6, and substitutes the value of the data item in the region characteristic data 110 into D_j . In this manner, the data complementing unit obtains y being the complement data information 270.

[0101]

Fig. 15 illustrates an example of the use complement model selection screen 70 displayed by the user terminal 20 in s15 of Fig. 5. As illustrated in Fig. 15, the use complement model selection screen 70 includes a complement data item display field 71, a recommended use complement model information display field 72, a significance ratio display field 76, a use model number registration field 77, a complement data correction registration field 78, and a "next" button 79.

[0102]

In the complement data item display field 71, the complement data item in the user setting information 210 is displayed.

[0103]

In the recommended use complement model information display field 72, information regarding the complement model selected by the complement model selection unit 107 from the complement models for all the explanatory variable patterns (for example, complement model with the minimum deviation) is set. As illustrated in Fig. 15, the recommended use complement model information display field 72 includes an explanatory variable pattern display field 73, a region-unit conversion value display field 74, and a deviation display field 75. Among the fields, the pattern number of the explanatory variable pattern used to generate the recommended complement model is displayed in the explanatory variable pattern display field 73. The total amount of the cell-unit complement data for each generated cell region is displayed in the region-unit conversion value display field 74. In the deviation display field 75, the value of the difference obtained from Equation 1 described above is displayed.

[0104]

The significance ratio display field 76 includes a display field for the pattern number of the explanatory variable pattern used to generate the complement model, a display field for the item of the explanatory variable, and a display field for the significance ratio of whether or not each complement model is significant when regression analysis

is performed. The display field of the significance ratio includes a significance ratio display field for cell analysis and a significance ratio display field for region analysis.

[0105]

In a use model number registration field 77, the user designates the complement model to be used by the pattern number of the explanatory variable.

[0106]

The complement data correction registration field 78 includes a checkbox for causing the user to designate whether or not the cell-unit complement data generated by the use complement model corresponding to the pattern number designated by the use model number registration field 77 is corrected, when the complement data item is included in the region characteristic data.

[0107]

When the next button 79 is operated, the user terminal 20 transmits the pattern number designated in the use model number registration field 77 and the information regarding the necessity of correction, which is input in the complement data correction registration field 78, to the data complementing apparatus 100 (s16).

[0108]

Fig. 16 illustrates an example of the complement information confirmation screen 80 displayed by the user

terminal 20 in s17 of Fig. 5. As illustrated in Fig. 16, the complement information confirmation screen 80 includes a complement data item display field 81, a cell-unit complement data information display field 82, a use complement model information display field 83, and the like.

[0109]

The complement data items are displayed in the complement data item display field 81, and cell-unit complement data based on the use complement model is displayed in the cell-unit complement data information display field 82 for each cell region.

[0110]

Information regarding the use complement model (explanatory variable 84 used to generate the use complement model, reason 85 for selecting the use complement model, and the like) is displayed in the use complement model information display field 83. The reason 85 for selecting the use complement model is a field for causing the user to select information such as, for example, that the deviation is small or the model has high significance.

[0111]

When the pattern number of the explanatory variable pattern other than the explanatory variable patterns displayed in the recommended use complement model information display field 72 on the use complement model selection screen

70 is input to the use model number registration field 77 by the user terminal 20, the data complementing apparatus 100 may display this as the reason 85 for which the use complement model is selected.

[0112]

Fig. 17 illustrates an example of the complement model information 260 in which the information of the complement model generated in the complement model generation process s25 is stored. The complement model information 260 includes information in which the complement model equation, the region conversion value, the deviation, the significance ratio for the cell analysis and the region analysis is associated with each explanatory variable pattern (pattern number). The complement model information 260 is used as information displayed on the use complement model selection screen 70.

[0113]

Fig. 18 illustrates an example of the complement data information 270 in which information regarding the complement data generated by the complement data generation process s5 is stored. As illustrated in Fig. 18, the complement data information 270 includes information in which the pattern number of the explanatory variable pattern (p01, p02, p03,...) and the complement data (for example, value of the number of owned vehicles) are associated with each other for

each of cell regions (U-001, U-002,...) belonging to a region U. For example, the illustrated complement data information 270 includes information indicating that the complement data calculated by using the complement model generated by the p01 explanatory variable pattern is 780 in the cell region corresponding to U-001.

[0114]

The complement data of the complement data information 270 is corrected by the complement data correction unit 108 when an instruction to perform correction is issued in the complement data correction registration field 78 of the use complement model selection screen 70 illustrated in Fig. 15 (S29 in Fig. 5). The above correction is performed by, for example, using the ratio of the deviation between the region conversion value and the value of the actual complement data item in the region characteristic data as a coefficient, and multiplying the coefficient by the complement data for each cell region. The complement data correction unit 108 updates the complement data information 270 according to the correction content.

[0115]

<Effect>

According to the data complementing apparatus 100 of the embodiment described above, it is possible to efficiently

supplement the missing data in the data related to the region by using the existing information.

[0116]

The data complementing apparatus 100 generates a complement model that complements the value of the missing data item in the cell region in which the region is divided into a mesh based on the external region characteristic data 230 of the external region, and the complement data is based on the generated complement model. Therefore, complement data can be generated accurately even though there is no external region similar to the region.

[0117]

Further, since the data complementing apparatus 100 tests the significance of the regression equation using the significance level received from the user, the user can efficiently obtain the complement data with the required accuracy according to the purpose of the service and the like.

[0118]

Since the data complementing apparatus 100 generates a complement model for each of the plurality of explanatory variable patterns, the user can select an appropriate complement model from the plurality of complement data generated based on the complement models having different explanatory variable patterns.

[0119]

Since the data complementing apparatus 100 generates the complement data based on the complement model having the highest reliability based on a plurality of complement models, the user can obtain the optimum complement data without comparing and examining a plurality of pieces of complement data.

[0120]

The data complementing apparatus 100 can appropriately correct the complement data based on a deviation when there is the deviation between the aggregated value of the complement data of the cell region of the region and the value of the data item corresponding to the missing data item in the region characteristic data.

[0121]

It should be noted that the embodiments described above are merely examples for explaining the present invention in an easy-to-understand manner, and the concept of the present invention includes various modifications and applications to the above embodiments. Those skilled in the art will recognize that various changes and modifications may be made in form and detail without departing from the spirit and scope of the claimed subject matter.

[0122]

For example, some components in the embodiment can be replaced with the components in another embodiment, and the configuration of another embodiment can be added to the configuration of the embodiment. Regarding some components in the embodiments, other components can be added, deleted, and replaced.

[0123]

For example, the cell region 301 is not limited to a rectangular region, and may have another shape such as an equilateral triangle as long as the region 300 is divided into predetermined regions having the same shape without gaps.

[0124]

Further, for example, the data complementing apparatus 100 may be configured to also function as the user terminal 20.

[0125]

In the above embodiment, the data complementing apparatus 100 generates a complement model for calculating cell-unit complement data by using multiple regression analysis, but the complement model may be generated using other analysis methods such as quantification analysis and cluster analysis.

[0126]

In the above-described embodiment, control lines and

information lines considered necessary for the descriptions are illustrated, and not all the control lines and the information lines in the product are necessarily shown. All components may be connected to each other. Further, in the above description, various types of information are illustrated by a table form (table), but the pieces of information may be managed in a form other than the table.

What is claimed is:

1. A data complementing system including an information processing apparatus, the system comprising:

a storage unit configured to store

region characteristic data that includes values of a plurality of data items regarding a predetermined region,

cell-region characteristic data that includes values of a plurality of data items regarding a cell region that is a region obtained by dividing the region into a mesh,

information indicating a missing data item that is the data item of missing data being data missed in the cell-region characteristic data,

external region characteristic data that includes values of a plurality of data items regarding an external region that is different from the region, and

an external cell-region characteristic data that includes values of a plurality of data items regarding an external cell region obtained by dividing the external region into a mesh; and

a complement model generation unit configured to generate a complement model for generating complement data that is for complementing the missing data, based on the external region characteristic data and the external cell-region characteristic data,

wherein

the complement model generation unit is configured to generate a first regression equation for each of a plurality of the external regions, the first regression equation in which the missing data item is expressed with the data item in the external cell-region characteristic data, which corresponds to the missing data item,

generate a second regression equation for all combinations obtained by selecting two of the plurality of the external regions, the second regression equation in which a difference between partial regression coefficients in the first regression equation for each of the external regions is used as an objective variable, and a difference between the values of the same data items in the external region characteristic data for the two selected external regions is used as an explanatory variable, and

generate the complement model based on the first regression equation and the second regression equation, the complement model in which the missing data item is expressed with the data item included in an explanatory variable pattern, the partial regression coefficient in the first regression equation, and a partial regression coefficient in the second regression equation, the explanatory variable pattern being a combination of one or more data items in the cell-region characteristic data,

wherein

the storage unit is configured to store a plurality of the explanatory variable patterns, and

the complement model generation unit is configured to generate the complement model for each of the plurality of the explanatory variable patterns.

2. The data complementing system according to claim 1, wherein

the storage unit is configured to store a significance level used for determining significance of each of a plurality of the complement models, and

the complement model generation unit is configured to determine the significance of the first regression equation and the significance of the second regression equation based on the significance level for each of the plurality of the complement models, and

generate a first significance determination result indicating the significance of the first regression equation and a second significance determination result indicating the significance of the second regression equation, for each of the plurality of the complement models.

3. The data complementing system according to claim 2, further comprising:

a complement model selection unit configured to obtain a deviation between the value of the data item corresponding to the missing data item in the region characteristic data and a sum of the values of the complement data for all cell regions belonging to the predetermined region, for each of the plurality of the complement models, output at least any of the deviation, the first significance determination result, and the second significance determination result, and receive a selection of the complement model corresponding to each of the plurality of the explanatory variable patterns.

4. The data complementing system according to claim 3, further comprising:

a data complementing unit configured to generate the complement data based on the selected complement model.

5. The data complementing system according to claim 4, further comprising:

a complement data correction unit configured to correct the generated complement data based on the deviation.

6. The data complementing system according to claim 4, further comprising:

an output unit configured to output at least any of information on the selected complement model and complement data generated by the complement model.

7. The data complementing system according to claim 1, further comprising:

a receiving unit configured to receive an input of at least any of the region characteristic data, the cell-region characteristic data, and the information indicating the missing data item.

8. A data complementing method, the method comprising:
by an information processing apparatus,
storing

region characteristic data that includes values of a plurality of data items regarding a predetermined region,

cell-region characteristic data that includes values of a plurality of data items regarding a cell region that is a region obtained by dividing the region into a mesh,

information indicating a missing data item that is the data item of missing data being data missed in the cell-region characteristic data,

external region characteristic data that includes values of a plurality of data items regarding an external region that is different from the region, and

an external cell-region characteristic data that includes values of a plurality of data items regarding an external cell region obtained by dividing the external region into a mesh; and

generating a complement model for generating complement data that is for complementing the missing data, based on the external region characteristic data and the external cell-region characteristic data,

wherein the data complementing method further comprises:

by the information processing apparatus,

generating a first regression equation for each of a plurality of the external regions, the first regression equation in which the missing data item is expressed with the data item in the external cell-region characteristic data, which corresponds to the missing data item;

generating a second regression equation for all combinations obtained by selecting two of the plurality of the external regions, the second regression equation in which a difference between partial regression coefficients in the first regression equation for each of the external regions is used as an objective variable, and a difference between the values of the same data items in the external region characteristic data for the two selected external regions is used as an explanatory variable; and

generating the complement model based on the first regression equation and the second regression equation, the complement model in which the missing data item is expressed with the data item included in an explanatory variable pattern, the partial regression coefficient in the first regression equation, and a partial regression coefficient in the second regression equation, the explanatory variable pattern being a combination of one or more data items in the cell-region characteristic data,

wherein the data complementing method further comprises:

by the information processing apparatus,

storing a plurality of the explanatory variable patterns; and

generating the complement model for each of the plurality of the explanatory variable patterns.

9. The data complementing method according to claim 8, further comprising:

by the information processing apparatus,

storing a significance level used for determining significance of each of a plurality of the complement models;

determining the significance of the first regression equation and the significance of the second regression equation based on the significance level for each of the

plurality of the complement models; and

generating a first significance determination result indicating the significance of the first regression equation and a second significance determination result indicating the significance of the second regression equation, for each of the plurality of the complement models.

10. The data complementing method according to claim 9, further comprising:

by the information processing apparatus,

obtaining a deviation between the value of the data item corresponding to the missing data item in the region characteristic data and a sum of the values of the complement data for all cell regions belonging to the predetermined region, for each of the plurality of the complement models;

outputting at least any of the deviation, the first significance determination result, and the second significance determination result; and

receiving a selection of the complement model corresponding to each of the plurality of the explanatory variable patterns.

11. The data complementing method according to claim 10, further comprising:

by the information processing apparatus,

generating the complement data based on the selected complement model.