

S P E C I F I C A T I O N

[TITLE]

IMPROVED PREDICTION SYSTEM AND PREDICTION METHOD USING
HETEROGENEOUS DATA

[TECHNICAL FIELD]

The present disclosure relates to improved prediction systems and prediction methods using heterogeneous data. More particularly, the present disclosure relates to an improved prediction system and an improved prediction method each using heterogeneous data, in which one or more prediction models are generated via artificial intelligence (AI) learning, based on one or more pieces of heterogeneous data including different features and basic data, and prediction is performed using a prediction model having highest prediction performance among the one or more prediction models.

[BACKGROUND ART]

Artificial intelligence (AI) learning, called machine learning, builds an algorithm that performs learning and prediction from learning data without additional programming. By constructing a prediction model that performs accurate prediction, a business operator can gain core competitiveness that differentiates itself from other competitors. However, since an external environment is constantly changing, prediction accuracy of a previously constructed prediction model gradually deteriorates due to external factors, and a new phenomenon cannot be predicted.

For example, in the case of a fraud prediction scenario used in the retail industry, since new types of fraud continue to appear, an old prediction model cannot detect new fraud types.

[Prior Art Documents]

[Patent Document]

(Patent document 0001) KR 10-1851675 B1 (May 17. 2018)

[DESCRIPTION OF EMBODIMENTS]

[TECHNICAL PROBLEM]

Provided is a prediction system and method constructing a prediction model capable of coping with new cases or new external factors by updating a previously-constructed prediction model.

Provided is a prediction system and method having improved prediction accuracy by generating prediction models by using data generated in various fields and making the generated prediction models compete with each other.

Technical problems of the present disclosure are not limited to the contents mentioned above, and other effects not mentioned will be clearly understood by one of ordinary skill in the art from the following description.

[SOLUTION TO PROBLEM]

According to an aspect of the present disclosure, an improved prediction system using the heterogeneous data includes a prediction model generator configured to generate one or more prediction models via artificial intelligence (AI) learning, based on basic data and one or more pieces of heterogeneous data including different features from the feature of the basic data, and a champion model selector configured to select a prediction model having highest prediction accuracy as a champion model by analyzing performance of each of the one or more prediction models, wherein prediction is performed using the selected champion model.

The champion model selector may re-select the champion model according to a preset condition.

The preset condition may be degradation of prediction accuracy of the champion model compared with prediction accuracy of another prediction model.

The preset condition may be degradation of prediction accuracy of the champion model compared with a predetermined reference value.

The prediction model generator may generate a new prediction model by using new heterogeneous data including a new feature different from the feature of existing data, during re-selection of the champion model, and the champion model selector may re-select the champion model from among the existing prediction model and the new prediction model.

The improved prediction system using the heterogeneous data may further include a major feature selector configured to select a major feature affecting prediction

accuracy of each of the one or more prediction models, based on a weight of the feature of each of the one or more prediction models.

The prediction model generator may generate a new prediction model via AI learning, based on the major feature selected by the major feature selector, and the champion model selector may select the champion model by including the new prediction model as the champion model candidate.

The heterogeneous data may be received from one or more external servers.

According to an embodiment of the present disclosure, an improved prediction method using the heterogeneous data includes a prediction model generation operation of generating one or more prediction models via AI learning, based on basic data and one or more pieces of heterogeneous data including different features from the feature of the basic data, and a champion model selection operation of selecting a prediction model having highest prediction accuracy as a champion model by analyzing performance of each of the one or more prediction models, and a prediction operation of performing prediction using the selected champion model.

The champion model selection operation may include re-selecting the champion model according to a preset condition.

The preset condition may be degradation of prediction accuracy of the champion model compared with prediction accuracy of another prediction model.

The preset condition may be degradation of prediction accuracy of the champion model compared with a predetermined reference value.

The prediction model generation operation may include generating a new prediction model by using new heterogeneous data including a new feature different from the feature of existing data, during re-selection of the champion model, and the champion model selection operation may include re-selecting the champion model from among the existing prediction model and the new prediction model.

The improved prediction method may further include a weight calculation operation of calculating a weight of the feature of each of the one or more prediction models, an influence determination operation of determining a level of influence upon the prediction accuracy of each of the prediction models, based on the weights, and a major feature selection operation of selecting a major feature from the features, based on the influences.

The prediction model generation operation may include generating a new prediction model via AI learning, based on the major feature selected in the major feature selection operation, and the champion model selection operation may include selecting the champion model by including the new prediction model as the champion model candidate.

[ADVANTAGEOUS EFFECTS OF DISCLOSURE]

In an improved prediction system and prediction method using heterogeneous data according to an embodiment of the present disclosure, a previously-constructed prediction model is periodically updated, and thus high prediction accuracy may be maintained even when new cases or new external factors appear.

In addition, prediction models are generated using data generated in various fields and compete with each other, and thus prediction accuracy may improve.

Moreover, the improved prediction system and prediction method operate with a continuous process in accordance with a predetermined condition or a schedule, and thus a periodic model re-construction time may be saved.

Effects of the present disclosure are not limited to the contents mentioned above, and other effects not mentioned will be clearly understood by one of ordinary skill in the art from the following description.

[BRIEF DESCRIPTION OF DRAWINGS]

FIG. 1 is a block diagram of an improved prediction system using heterogeneous data, according to an embodiment of the present disclosure.

FIG. 2 illustrates a structure in which an improved prediction system using heterogeneous data according to an embodiment of the present disclosure employs a champion model.

FIG. 3 is a block diagram of an improved prediction system using heterogeneous data, according to an embodiment of the present disclosure.

FIG. 4 is a table showing an equation and predication accuracy for each prediction model, according to an embodiment of the present disclosure.

FIG. 5 is a block diagram illustrating a champion model re-selection process by an improved prediction system using heterogeneous data, according to an embodiment

of the present disclosure.

FIG. 6 is a flowchart of an improved prediction method using heterogeneous data, according to an embodiment of the present disclosure.

FIG. 7 is a flowchart of prediction model updating for improving prediction accuracy, according to an embodiment of the present disclosure.

FIG. 8 is a flowchart of an improved prediction method using heterogeneous data including a major feature selection operation, according to an embodiment of the present disclosure.

[DETAILED DESCRIPTION OF DISCLOSURE]

Embodiments of the disclosure are described in detail herein with reference to the accompanying drawings so that this disclosure may be easily performed by one of ordinary skill in the art to which the disclosure pertains. The disclosure may, however, be embodied in many different forms and should not be construed as being limited to the embodiments set forth herein. In the drawings, parts irrelevant to the description are omitted for simplicity of explanation.

The terms used in the disclosure are merely used to describe particular embodiments, and are not intended to limit the scope of the disclosure. An expression used in the singular may encompass the expression of the plural, unless it has a clearly different meaning in the context.

In the present specification, it is to be understood that the terms such as "including," "having," and "comprising" are intended to indicate the existence of the features, numbers, steps, actions, components, parts, or combinations thereof disclosed in the specification, and are not intended to preclude the possibility that one or more other features, numbers, steps, actions, components, parts, or combinations thereof may exist or may be added.

In addition, the components shown in the embodiments of the present disclosure are shown independently to indicate different characteristic functions, and thus do not mean that each component is composed of separate hardware or one software component unit. In other words, for convenience of description, each component is listed as each component, and at least two components may be combined to form one component, or one component may be divided into a plurality of components to perform

a function. An embodiment in which at least two components may be combined to form one component, and an embodiment in which one component is divided into a plurality of components are also included in the scope of the present disclosure without departing from the essence of the present disclosure.

The following embodiments are provided to be more clearly explained to those of ordinary skill in the art, and thus the shapes and sizes of elements in the drawings may be exaggerated for more clear explanation.

Embodiments of the present disclosure will now be described more fully with reference to the accompanying drawings.

FIG. 1 is a block diagram of an improved prediction system using heterogeneous data according to an embodiment of the present disclosure, and FIG. 2 illustrates a structure in which the improved prediction system using the heterogeneous data according to an embodiment of the present disclosure employs a champion model.

Referring to FIGS. 1 and 2, an improved prediction system 10 using heterogeneous data according to an embodiment of the present disclosure may include a prediction model generator 100, a champion model selector 200, a control processor 300, and a memory 400. The components of FIG. 1 are not essential to implement the prediction system 10 using the heterogeneous data, and the improved prediction system 10 using heterogeneous data described herein may include more or less components than the components of FIG. 1.

The prediction model generator 100 may generate one or more prediction models via artificial intelligence (AI) learning, based on basic data 21, and one or more pieces of heterogeneous data 22 and 23 including different features from the feature of the basic data 21.

According to the present embodiment, it is described supposing that data 20 is structured data. However, the data 20 may not only include the structured data but also may include unstructured data and semi-structured data.

According to the present embodiment, the basic data 21 may be data directly generated or held by a subject that generates a prediction model, and may refer to data used as learning data to generate the prediction model. For example, conventionally, in order to generate a prediction model that predicts the default rate of credit loan customers in a financial institution such as a bank, the financial institution uses

customer data generated by the financial institution itself (customer's age, residence, gender, income, workplace, work period, asset status information, credit information, overdue information, repayment information, and product information) as learning data. Here, the customer data generated by the financial institution itself may be the basic data 21.

According to the present embodiment, the heterogeneous data 22 and 23 may refer to data not directly generated or not held by the subject generating the prediction model. For example, when a financial institution is the subject generating the prediction model, the heterogeneous data 22 and 23 may refer to customer data of an airline other than the financial institution, customer communication record data of a telecommunication company, and customer purchase data of an e-commerce company.

In more detail, the heterogeneous data 22 and 23 may refer to data including at least one new feature other than the features included in the basic data 21. For example, when the basic data 21 includes features A, B, and C, data including features A, B, and D or features A, B, C, and D may be referred to as the heterogeneous data 22 and 23 different from the basic data 21.

A feature means a characteristic, feature, or attribute capable of defining a difference between data values included in data, and is referred to as a variable, a field, a dimension, or the like. For example, when the data is financial customer data generated by the financial institution, the feature may mean a customer's age, residence, gender, income, workplace, working period, asset status information, credit information, overdue information, repayment information, product information, or the like.

The heterogeneous data 22 and 23 may be received from one or more external servers. For example, customer data of an airline may be received from an external server of the airline, customer communication record data of a telecommunication company may be received from an external server of the telecommunication company, and customer purchase data of an e-commerce company may be received from an external server of the e-commerce company.

The prediction model generator 100 may generate one or more prediction models. The prediction model generator 100 may generate the one or more prediction models by using different learning data. The different learning data may be

combinations of the heterogeneous data 22 and 23 based on the basic data 21. For example, the prediction model generator 100 may generate a first prediction model M_a , based on the basic data 21, generate a second prediction model M_{a+b} , based on data corresponding to a combination of the basic data 21 and data B (22), and generate a third prediction model M_{a+c} , based on data corresponding to a combination of the basic data 21 and data C (23). Although not shown in FIGS. 1 and 2, the prediction model generator 100 may generate prediction models by combining the basic data 21 with the data B (22), the data C (23), and data D (not shown).

A prediction model constructed using a combination of the basic data 21 with the heterogeneous data 22 and 23 as learning data may have higher prediction accuracy or higher prediction performance than a prediction model constructed using data of each field as learning data. For example, when only data of a search engine is used as learning data, prediction of a consumer's purchase interest may be accurate, but purchase prediction of a product may be less accurate in practice. On the other hand, when only card payment data is used as learning data, prediction of a consumer's product purchase may be accurate, but the accuracy of predicting a consumer's potential purchase tendency may deteriorate. In this case, a prediction model constructed using a combination of search engine data and card payment data as learning data enables highly accurate predictions of consumer's purchase tendency and potential purchase tendency.

The prediction model generator 100 may generate one or more prediction models by using a learning algorithm such as a Logistic Regression Support, a Support Vector Machine, a Deep Neural Network, a Convolutional Neural Network, a Kth-Nearest Neighbor, a Decision Tree, a Random Forest, eXtreme Gradient Boosting (XGBoost), Categorical Gradient Boosting (CatBoost), AdaBoosting (AdaBoost), Gradient Boosting, Light Gradient Boosting, or Extra Trees Boosting, according to a target value set based on a plurality of learning data obtained via combinations.

The prediction model generator 100 may train or generate the one or more prediction models to output the same target value.

It may be difficult to determine what combination of combination data has higher prediction performance from among combinations of the basic data 21 and the one or more heterogeneous data 22 and 23, before direct prediction is performed using

prediction models constructed with the pieces of combination data. Accordingly, there is a need to determine a prediction model showing highest performance by configuring various heterogeneous data combinations and performing a performance evaluation with respect to prediction models respectively constructed with pieces of data corresponding to the various combinations.

The champion model selector 200 may select a prediction model having highest prediction accuracy as a champion model by analyzing the performances of the one or more prediction models generated by the prediction model generator 100. First, the champion model selector 200 may calculate performance indexes for the one or more prediction models, and then select and designate, as a champion model M_{a+b} , one prediction model M_{a+b} determined to have best performance via a comparison between the calculated performance indexes.

According to a learning algorithm used in the prediction model generation, the performance indexes may be calculated using at least one of a Receiver Operating Characteristic Curve _ Area Under the Curve (ROC_AUC), a Precision-Recall Curve _ Area Under the Curve (PRC_AUC), a Log Loss, a Root Mean Square Error (RMSE), a Mean Absolute Error (MAE), or a determination coefficient (R-Square).

The control processor 300 may control the prediction model generator 100 and the champion model selector 200 to generate one or more prediction models M_a , M_{a+b} , and M_{a+c} , based on each learning data, and to evaluate performances for the generated one or more prediction models M_a , M_{a+b} , and M_{a+c} and select a champion model M_{a+b} .

The control processor 300 may perform prediction to obtain a target value that is to be predicted using the selected champion model M_{a+b} selected by the champion model selector 200. For example, the control processor 300 may perform prediction to input new vehicle insurance subscriber data to the champion model M_{a+b} selected by the champion model selector 200 with respect to prediction of the possibility of occurrence of an illegal vehicle insurance claim by a customer and to calculate the possibility that a new subscriber makes an illegal claim.

The control processor 300 may continuously monitor all prediction models generated to determine performance of the champion model. For example, in the case of a fraud detection model, the performance of a current champion model of which initial data characteristics change is highly likely to degrade over time. The control processor

300 may ascertain the performance of the current champion model by continuously monitoring all prediction models, thereby determining a replacement time point of the current champion model.

The memory 400 may store one or more programs for operations of the prediction model generator 100, the champion model selector 200, and the control processor 300, and may also store various types of data (basic data, heterogeneous data, a prediction model input value, etc.) necessary to perform the operations of the prediction model generator 100, the champion model selector 200, and the control processor 300.

The improved prediction system 10 using the heterogeneous data according to an embodiment of the present disclosure may perform prediction by generating the one or more prediction models M_a , M_{a+b} , and M_{a+c} via combinations of the basic data 21 and the heterogeneous data 22 and 23 as described above and designating a prediction model determined to have best prediction performance via a performance evaluation as the champion model M_{a+b} .

FIG. 3 is a block diagram of an improved prediction system using heterogeneous data according to an embodiment of the present disclosure, and FIG. 4 is a table showing an equation and predication accuracy for each prediction model, according to an embodiment of the present disclosure.

Referring to FIGS. 3 and 4, the improved prediction system 10 using heterogeneous data according to an embodiment of the present disclosure may further include a major feature selector 500.

The major feature selector 500 may calculate a weight for the feature of each of the one or more prediction models generated by the prediction model generator 100, and may select a major feature affecting prediction accuracy, based on the calculated weights.

A combination of basic data and heterogeneous data is used as learning data to improve the prediction accuracy of a prediction model, but all features included in the heterogeneous data may not contribute to the prediction accuracy. In this case, it may be more efficient to select a major feature having a largest influence upon prediction accuracy and using the selected major feature as the learning data.

For example, as shown in FIG. 4, when the prediction model generator 100

generates prediction models M_A , M_{A+B} , M_{A+C} and M_{A+B+C} by using basic data A (including features A_1 and A_2), heterogeneous data B (including features B_1 and B_2), and heterogeneous data C (including features C_1 and C_2), the champion model selector 200 may calculate prediction accuracy of each of the prediction models M_A , M_{A+B} , M_{A+C} and M_{A+B+C} , and the major feature selector 500 may calculate respective weights w_1 , w_2 , w_3 , w_4 , w_5 , and w_6 for the prediction models M_A , M_{A+B} , M_{A+C} and M_{A+B+C} . For example, for the prediction model M_{A+B+C} of FIG. 4, when the major feature selector 500 calculates w_1 to be 0.1, w_2 to be 0.2, w_3 to be 0.3, w_4 to be 0.02, w_5 to be 0.2, w_6 to be 0.001, as the respective weights of the features, the major feature selector 500 may determine feature B_2 of the heterogeneous data B and feature C_2 of the heterogeneous data C in the prediction model M_{A+B+C} as a feature not contributing to improving the prediction accuracy of the prediction model M_{A+B+C} , and may determine feature B_1 of the heterogeneous data B and feature C_1 of the heterogeneous data C as major features that greatly affect improvement in the prediction accuracy of the prediction model M_{A+B+C} .

As such, when a major feature is determined from the heterogeneous data, the prediction model generator 100 may generate another prediction model by using only the major feature as the learning data, and the champion model selector 200 may use the prediction model trained with the major feature as a champion model candidate and select a champion model through a comparison between the prediction model and another prediction model.

According to the present embodiment, the heterogeneous data is received from an external server, and thus, to receive the heterogeneous data from the external server, a subject that generates a prediction model may transmit the heterogeneous data through a data purchase contract. In this case, heterogeneous data purchasing costs may be saved by determining a major feature. For example, as shown in FIG. 4, the prediction models M_{A+B} (95%), M_{A+C} (99%), and M_{A+B+C} (99.9%) using combinations of basic data with heterogeneous data as learning data may have better prediction accuracy than the prediction model M_A (80%) using only the basic data as the learning data, and the prediction model M_{A+B+C} among the prediction models M_{A+B} (95%), M_{A+C} (99%), and M_{A+B+C} (99.9%) may provide best prediction accuracy. However, in this case, all of a plurality of heterogeneous data need to be purchased, and thus there may

be a burden in terms of costs. When a major feature is selected via a comparison between the weights of features B_1 and B_2 in the heterogeneous data B and the heterogeneous data B is selectively purchased, data purchase costs may be saved.

FIG. 5 is a block diagram illustrating a champion model re-selection process by an improved prediction system using heterogeneous data, according to an embodiment of the present disclosure.

Referring to FIGS. 1, 2 and 5, the selected champion model M_{a+b} provides good performance at the time point when performance evaluation has been performed, but, because several external factors change at a future moment, prediction performance may deteriorate. To prevent such deterioration of prediction performance due to a change in external factors over time and further improve prediction performance, the champion model selector 200 may re-select a champion model according to a preset condition. The re-selection may be performed with respect to the one or more prediction models generated by the prediction model generator 100 by inputting an input value input at a time point when re-selection is performed by the champion model selector 200 to the one or more prediction models to calculate performance index values and comparing the calculated performance index values with each other.

The preset condition may be deterioration of prediction accuracy of the previously selected champion model M_{a+b} compared with prediction accuracy of the other prediction models M_a and M_{a+c} , or deterioration of prediction accuracy of the champion model M_{a+b} compared with a predetermined reference value.

Referring to FIG. 5, the improved prediction system 10 using heterogeneous data according to an embodiment of the present disclosure may generate prediction models other than the previously generated prediction models M_a , M_{a+b} , and M_{a+c} and use the other prediction models as champion model candidates, during champion model re-selection. In more detail, the control processor 300 may remove the prediction model M_{a+c} providing lowest performance in the performance evaluation at a time point of re-selecting a champion model, from a champion model candidate group. The prediction model generator 100 may generate a new prediction model M_{a+d} by using heterogeneous data including other new features other than previously-used data, and the champion model selector 200 may designate the new prediction model M_{a+d} as a challenger model and may put the new prediction model M_{a+d} into the champion model

candidate group. Thereafter, the champion model selector 200 may perform performance evaluation by including the challenger model M_{a+d} newly put into the champion model candidate group, and may newly select the prediction model M_{a+d} determined to provide highest performance according to the performance index as a champion model.

According to the present embodiment, the challenger model has been described above as being constructed using new heterogeneous data as learning data. However, embodiments are not limited thereto. The challenger model may be a prediction model constructed with latest data updated from the previously-used learning data as learning data, or may be a prediction model constructed by changing a learning algorithm while using the previously-used learning data without changes.

The improved prediction system 10 using heterogeneous data according to an embodiment of the present disclosure may enable a series of processes such as continuous monitoring, testing, re-education, and re-distribution for coping with a currently constantly changing condition to be automatically performed without depending on human hands, and may also enable improvement or maintenance of prediction performance of the entire prediction system by updating or replacing a prediction model according to a preset condition or a preset cycle.

FIG. 6 is a flowchart of an improved prediction method using heterogeneous data, according to an embodiment of the present disclosure.

Referring to FIG. 6, the improved prediction method using heterogeneous data, according to an embodiment of the present disclosure, may include a prediction model generation operation S110, a champion model selection operation S120, and a prediction operation S130.

In the prediction model generation operation S110, one or more prediction models may be generated via AI learning, based on basic data and one or more pieces of heterogeneous data including different features from the feature of the basic data.

In the champion model selection operation S120, a prediction model having highest prediction accuracy may be selected as a champion model by analyzing the performances of the one or more prediction models.

In the prediction operation S130, prediction may be performed using the champion model.

FIG. 7 is a flowchart of prediction model updating for improving prediction accuracy, according to an embodiment of the present disclosure.

In the improved prediction method using heterogeneous data according to an embodiment of the present disclosure, to determine prediction accuracy of the champion model, performance may be monitored by continuing the generated one or more prediction models, in operation S210. The monitoring operation S210 is performed to detect degradation of prediction performance of the champion model that is used for the prediction. When it is determined in operation S220 that degradation of prediction performance of the champion model is not detected, prediction may be performed using the existing champion model. On the other hand, when it is determined in operation S220 that degradation of prediction performance of the champion model is detected, one or more new prediction models, namely, one or more challenger models, replacing the existing champion model may be generated.

The challenger model may be a prediction model constructed using new heterogeneous data as learning data, a prediction model newly trained by updating the previously-used learning data with latest data, or a prediction model constructed by changing a learning algorithm while using the previously-used learning data without changes.

When the one or more the challenger models are generated, a prediction model providing best performance via a comparison with the performance of the existing champion model may be selected as a new champion model, in operation S240. When the new champion model is selected, prediction may be performed using the newly selected champion model, in operation S250.

The monitoring operation S210, the challenger model generation operation S230, the champion model re-selection operation S240, and the prediction operation S250 are not performed once, but may be continuously repeated.

In the improved prediction method using heterogeneous data according to an embodiment of the present disclosure, a prediction model (champion model) for use in prediction is continuously updated, and thus prediction accuracy of a prediction model may be improved or maintained and a risk of decision making due to wrong prediction may be reduced.

FIG. 8 is a flowchart of an improved prediction method using heterogeneous

data including a major feature selection operation, according to an embodiment of the present disclosure.

Referring to FIG. 8, the improved prediction method using heterogeneous data, according to an embodiment of the present disclosure, may include a weight calculation operation S320, an influence determination operation S330, and a major feature selection operation S340, and may further include a second prediction model generation operation S350.

In the weight calculation operation S320, the weight of the feature of each of the prediction models may be calculated. In the influence determination operation S330, a level of influence upon the prediction accuracy of each of the prediction models may be determined based on the weights calculated in the weight calculation operation S320.

In the major feature selection operation S340, a major feature may be selected from the features, based on the influences determined in the influence determination operation S330. When the major feature is selected, the second prediction model generation operation S350 of generating another prediction model via AI learning by extracting the major feature from the features of the basic data and the heterogeneous data may be performed.

The heterogeneous data may be received from an external server, and only heterogeneous data including the major feature may be selected and received. In this case, purchase costs of the heterogeneous data may be saved due to the major feature determination.

When the other prediction model is generated in the second prediction model generation operation S350, the other prediction model may also be included as a champion model candidate and the champion model may be selected, in the champion model selection operation S360.

Various embodiments described in the specification may be implemented by hardware, middleware, a microcode, software and/or combinations thereof. For example, various embodiments may be implemented in one or more on-demand semiconductors (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), processors, controllers, microcontrollers, microprocessors, other electronic units designed to perform the functions presented herein, or combinations thereof.

For example, various embodiments may be recorded or encoded in a computer-readable medium including commands. The commands recorded or encoded on the computer-readable medium may cause a programmable processor or other processors to perform a method, for example, when the commands are executed. Computer-readable media include both computer storage media and communication media including any medium that facilitates transfer of a computer program from one place to another place. Storage media may be any available media that can be accessed by a computer. For example, such computer-readable media may include RAM, ROM, EEPROM, CD-ROM or other optical disk storage media, magnetic disk storage media, other magnetic storage devices, or any other medium that can be used to transport or store a desired program code in the form of structures commands or data structures accessible by a computer.

Such hardware, software, firmware, and the like may be implemented within the same device or within separate devices to support various operations and functions described herein. Additionally, the elements, units, modules, components, and the like described as "~er(or)" in the present disclosure may be implemented together or may be individually implemented as individual but interoperable logic devices. Depiction of different features for modules, units, etc. is intended to highlight different functional embodiments, and does not necessarily mean that they must be realized by individual hardware or software components. Rather, functions associated with one or more modules or units may be performed by separate hardware or software components or may be incorporated within common or separate hardware or software components.

Although operations are shown in the drawings in a specific order, it should not be understood that these operations are performed in the specific order or in a sequential order to achieve a desired result, or that all the illustrated operations should be performed. In any environment, multitasking and parallel processing may be advantageous. Moreover, division of various components in the above-described embodiments should not be understood as being needed in all embodiments, and the described components will generally be understood as being integrated together into a single software product or packaged into multiple software products.

While the present disclosure has been particularly shown and described with reference to exemplary embodiments thereof, it will be understood by one of ordinary

skill in the art that various changes in form and details may be made therein without departing from the spirit and scope as defined by the following claims.

[Description of reference numerals]

10: prediction system using heterogeneous data

100: prediction model generator

200: champion model selector

300: control processor

400: memory

500: major feature selector

CLAIMS

1. An improved prediction system using heterogeneous data, the improved prediction system comprising:
 - a prediction model generator configured to generate one or more prediction models via artificial intelligence (AI) learning, based on basic data and one or more pieces of heterogeneous data including different features from the feature of the basic data; and
 - a champion model selector configured to select a prediction model having highest prediction accuracy as a champion model by analyzing performance of each of the one or more prediction models,wherein prediction is performed using the selected champion model.
2. The improved prediction system of claim 1, wherein the champion model selector is further configured to re-select the champion model according to a preset condition.
3. The improved prediction system of claim 2, wherein the preset condition is degradation of prediction accuracy of the champion model compared with prediction accuracy of another prediction model.
4. The improved prediction system of claim 2, wherein the preset condition is degradation of prediction accuracy of the champion model compared with a predetermined reference value.
5. The improved prediction system of claim 2, wherein
 - the prediction model generator is further configured to generate a new prediction model by using new heterogeneous data including a new feature different from the feature of existing data, during re-selection of the champion model, and
 - the champion model selector is further configured to re-select the champion model from among the existing prediction model and the new prediction model.
6. The improved prediction system of claim 1, further comprising a major

feature selector configured to select a major feature affecting prediction accuracy of each of the one or more prediction models, based on a weight of the feature of each of the one or more prediction models.

7. The improved prediction system of claim 6, wherein
the prediction model generator is further configured to generate a new prediction model via AI learning, based on the major feature selected by the major feature selector, and
the champion model selector is further configured to select the champion model by including the new prediction model as a champion model candidate.

8. The improved prediction system of claim 1, wherein the heterogeneous data is received from one or more external servers.

9. An improved prediction method using heterogeneous data, the improved prediction method comprising:

a prediction model generation operation of generating one or more prediction models via artificial intelligence (AI) learning, based on basic data and one or more pieces of heterogeneous data including different features from the feature of the basic data; and

a champion model selection operation of selecting a prediction model having highest prediction accuracy as a champion model by analyzing performance of each of the one or more prediction models; and

a prediction operation of performing prediction using the selected champion model.

10. The improved prediction method of claim 9, wherein the champion model selection operation comprises re-selecting the champion model according to a preset condition.

11. The improved prediction method of claim 10, wherein the preset condition is degradation of prediction accuracy of the champion model compared with

prediction accuracy of another prediction model.

12. The improved prediction method of claim 10, wherein the preset condition is degradation of prediction accuracy of the champion model compared with a predetermined reference value.

13. The improved prediction method of claim 10, wherein the prediction model generation operation comprises generating a new prediction model by using new heterogeneous data including a new feature different from the feature of existing data, during re-selection of the champion model, and the champion model selection operation comprises re-selecting the champion model from among the existing prediction model and the new prediction model.

14. The improved prediction method of claim 9, further comprising:
a weight calculation operation of calculating a weight of the feature of each of the one or more prediction models;
an influence determination operation of determining a level of influence upon the prediction accuracy of each of the prediction models, based on the weights; and
a major feature selection operation of selecting a major feature from the features, based on the influences.

15. The improved prediction method of claim 14, wherein the prediction model generation operation comprises generating a new prediction model via AI learning, based on the major feature selected in the major feature selection operation, and the champion model selection operation comprises selecting the champion model by including the new prediction model as a champion model candidate.