METHOD, APPARATUS, SERVER, AND USER TERMINAL FOR CONSTRUCTING DATA PROCESSING MODEL

Technical Field

[01] The present specification belongs to the field of Internet technologies, and in particular, to a method, an apparatus, a server, and a user terminal for constructing a data processing model.

Background

[02] With the development of artificial intelligence technologies, a data processing method that constructs a corresponding data processing model through machine learning, and then uses the constructed data processing model to perform specific data processing on target data is widely used in increasingly more fields.

[03] At present, when constructing the above data processing model based on an existing method, a user needs to simultaneously design and perform a test analysis and other operations (for example, including feature engineering, model selecting, model parameter searching, and so on) on a plurality of possible model analyses through a corresponding code program or the like to determine a suitable model that meets processing requirements of target data as a base model, and the base model is trained to construct a required target data processing model. During the implementation of the above method, the user is required to have relatively high professional capabilities (for example, testing of a model structure, data analysis on the model, and so on), and each time a data processing model is constructed, the user needs to manually perform a plurality of tests to find a suitable base model, which certainly will increase the complexity of the processing process and affect the processing efficiency. Therefore, a simpler and more efficient method for constructing a data processing model is in urgent need, so that it is unnecessary for a user to test and analyze tediously to search for a suitable base model, thereby reducing the operational difficulty of the user and improving the processing efficiency.

Summary of the Invention

[04] An objective of the present specification is to provide a method, an apparatus, a server, and a user terminal for constructing a data processing model, so that it is unnecessary for a user to test and analyze tediously to search for a suitable base model and construct a target data processing model, thereby reducing the operational difficulty of the user and improving the processing efficiency.

[05] The method, apparatus, server, and user terminal for constructing a data processing model provided in the present specification are implemented as follows.

[06] A method for constructing a data processing model, including: acquiring a model description parameter and sample data of a target data processing model; determining a base model according to the model description parameter and the sample data; and training the base model according to the sample data to obtain the target data processing model.

[07] A data processing method, including: acquiring to-be-processed target data; and invoking a target data processing model to process the target data, wherein the target data processing model is obtained by determining a base model according to a model description parameter and sample data of the target data processing model that are acquired and then training the base model according to the sample data.

[08] A data processing method, including: displaying a data input interface; and receiving a model description parameter and sample data of a target data processing model input by a user on the data input interface, wherein the model description parameter and the sample data are configured to determine a base model, and the base model is configured to establish the target data processing model.

[09] An apparatus for constructing a data processing model, including: an acquisition module configured to acquire a model description parameter and sample data of a target data processing model; a determination module configured to determine a base model according to the model

description parameter and the sample data; and a training module configured to train the base model according to the sample data to obtain the target data processing model.

[10] A server, including a processor and a memory configured to store processor-executable instructions, wherein when the processor executes the instructions, the following operations are implemented: acquiring a model description parameter and sample data of a target data processing model; determining a base model according to the model description parameter and the sample data; and training the base model according to the sample data to obtain the target data processing model.

[11] A user terminal, including a processor and a memory configured to store processor-executable instructions, wherein when the processor executes the instructions, the following operations are implemented: displaying a data input interface; and receiving a model description parameter and sample data of a target data processing model input by a user on the data input interface, wherein the model description parameter and the sample data are configured to determine a base model, and the base model is configured to establish the target data processing model.

[12] A computer-readable storage medium, storing computer instructions thereon, wherein when the instructions are executed, the following operations are implemented: acquiring a model description parameter and sample data of a target data processing model; determining a base model according to the model description parameter and the sample data; and training the base model according to the sample data to obtain the target data processing model.

[13] According to the method, apparatus, server, and user terminal for constructing a data processing model provided in the present specification, by acquiring a model description parameter in a targeted manner, a server can accurately determine a modeling requirement of a user based on the model description parameter, automatically match the modeling requirement of the user to a suitable model as a base model, and then construct a corresponding data processing model, so that it is unnecessary for the user to test and analyze tediously to determine a suitable base model, thereby reducing the operational difficulty of the user and improving the processing efficiency.

Brief Description of the Drawings

[14] To describe the technical solutions in the embodiments of the present application or the prior art more clearly, the accompanying drawings required for describing the embodiments or the prior art are briefly introduced below. It is apparent that the accompanying drawings described in the following are merely some embodiments of the present application, and those of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

[15] FIG. 1 is a schematic diagram of an embodiment of a system structure based on a method for constructing a data processing model according to an embodiment of the present specification;

[16] FIG. 2 is a schematic diagram of an embodiment of a data input interface displayed by using the method for constructing a data processing model according to an embodiment of the present specification in a scenario example;

[17] FIG. 3 is a schematic diagram of an embodiment in which candidate models are filtered by using the method for constructing a data processing model according to an embodiment of the present specification in a scenario example;

[18] FIG. 4 is a schematic diagram of another embodiment in which candidate models are filtered by using the method for constructing a data processing model according to an embodiment of the present specification in a scenario example;

[19] FIG. 5 is a schematic diagram of an embodiment of a process of the method for constructing a data processing model according to an embodiment of the present specification;

[20] FIG. 6 is a schematic diagram of an embodiment of a process of a data processing method according to an embodiment of the present specification;

[21] FIG. 7 is a schematic diagram of another embodiment of a process of the data processing method according to an embodiment of the present specification;

[22] FIG. 8 is a schematic diagram of an embodiment of a structure of a server according to an embodiment of the present specification;

[23] FIG. 9 is a schematic diagram of an embodiment of a structure of a user terminal according to an embodiment of the present specification; and

[24] FIG. 10 is a schematic diagram of an embodiment of a structure of an apparatus for constructing a data processing model according to an embodiment of the present specification.

Detailed Description

[25] In order to enable those skilled in the art to better understand the technical solutions in the present specification, the technical solutions in the embodiments of the present specification will be described clearly and completely with reference to the accompanying drawings in the embodiments of the present specification. It is apparent that the described embodiments are merely some of rather than all the embodiments of the present application. Based on the embodiments of the present specification, all other embodiments derived by those of ordinary skill in the art without any creative efforts shall fall within the protection scope of the present specification.

[26] Based on an existing method for constructing a data processing model, a user often needs to perform tests and analyses (which may specifically include, for example, feature engineering, model selecting, model parameter searching, and other steps) by coding or in other manner in the process of constructing a target data processing model through machine learning, and determine, from a variety of possible models, a suitable model as a base model (or referred to as an initial model). However, the above determining a suitable model as a base model has relatively high professional requirements for a user, and the operational difficulty is relatively high. For example, the user may be required to have high data analysis capabilities and programming knowledge in order to interpret and analyze underlying data (for example, program code and the like) fed back based on each test model to determine a degree of applicability of each test model to to-be-processed target data. In addition, since the user is required to find a suitable base model by himself/herself, the implementation process is relatively complicated and the processing efficiency is relatively low.

[27] For the root cause of the above problems, the present specification considers that the description of the model and the determination of a relevant base model can be separated. Specifically, on a user-oriented side, a user can be allowed to provide a relatively simple model description parameter to represent requirements and limitations of the user for a required base model, and then a platform server can automatically match the model description parameter to a suitable model as a base model for the user to enable the user to construct a required data processing model through training of the base model, so that it is unnecessary for the user to test and analyze tediously to determine a suitable base model, thereby reducing the operational difficulty of the user and improving the processing efficiency.

[28] Based on the above idea, a method for constructing a data processing model is provided in an implementation of the present specification. The method for constructing a data processing model is applicable to a system including a server and a user terminal. As shown in FIG. 1, the user terminal and the server are coupled to each other for specific data interaction.

[29] Based on the above system, a user may set, through the user terminal, a model description parameter that meets his/her own modeling requirement, and input corresponding sample data. The server may receive a model description parameter and sample data sent by the user terminal, and select a suitable model as a base model and a corresponding model parameter according to the model description parameter and the sample data set by the user, and then may provide the base model and the model parameter to the user, so that the user can directly use the base model to train and construct a target data processing model; and can also use the sample data input by the user to automatically train and construct the target data processing model based on the base model and the model parameter.

[30] In this implementation, the server may specifically be a background service server with data analysis, processing, and data communication functions. The server may be an electronic device with data computation, storage, and network interaction functions; or may be software running in the electronic device and providing support for data processing, storage, and network

interaction. In this implementation, the number of the server is not specifically limited. The server may be one or more servers, or a server cluster formed by several servers.

[31] In this implementation, the user terminal may be a port device with data input, data display, and data communication functions. Specifically, the user terminal may be, for example, a desktop computer, a tablet computer, a notebook computer, a smart phone, a digital assistant, a smart wearable device, a shopping guide terminal, a television with a network access function, or the like. Alternatively, the user terminal may also be a software application, a graphical interface, or the like that is runnable in the electronic device.

[32] In a specific scenario example, the user may apply the method for constructing a data processing model according to this embodiment of the present specification to easily and efficiently train and construct a target data processing model.

[33] In this scenario example, the user wants to construct a data processing model capable of identifying the gender of a character in a picture. If directly based on an existing method for constructing a data processing model, the user needs to test and analyze a variety of possible models by himself/herself to select a model suitable for identifying the gender of the character in the picture as a base model, and then train and construct a specific data processing model. For the user, the implementation is relatively tedious, the operation is difficult, and the processing efficiency is also relatively low.

[34] Based on the method for constructing a data processing model according to the embodiment of the present specification, a platform server may generate, based on a predefined model description rule, a data input interface for the user to input a model description parameter, and send the data input interface to the user terminal. The user terminal may display the above data input interface to the user. See FIG. 2 for details. A plurality of model description parameters to be set by the user are displayed on the data input interface.

[35] The above model description parameters may be specifically understood as parameter data that is set by the user and is capable of effectively representing data processing requirements and/or

user customization requirements of a target data processing model to be constructed. In this way, even if the user does not have enough understanding of the specific underlying data, or has relatively poor professional capabilities, the user can easily and efficiently set his/her own modeling requirement through the model description parameters according to data processing to be implemented by the target data processing model that the user intends to construct. Moreover, the above parameter data conforms to a model description rule. That is, the server may display the corresponding data input interface to the user through the user terminal based on the model description rule, so that the user can set specific model description parameters according to a format corresponding to the model description rule and then can accurately identify and determine specific information included in the above model description parameters. A model that meets implementation requirements and/or user customization requirements (that is, a suitable model) of the target data processing model to be constructed can be automatically found and matched for the user on the basis of the above information to serve as a base model.

[36] Specifically, the above model description parameters may include: a name of the target data processing model (which may be denoted as model_name), a data label (which may be denoted as label), a data feature (which may be denoted as feature), an evaluation metric (which may be denoted as eval_metrics), and the like. Of course, it should be noted that the model description parameters listed above are only for better illustrating the implementation of the present specification. During a specific implementation, according to a specific situation of the user and specific characteristics of the target data processing model to be constructed, other types of parameter data such as a model type may also be introduced as model description parameters. This is not limited in the present specification.

[37] The name of the target data processing model may be understood as an identification symbol of the data processing model that the user wants to construct, for example, an A model, a model for identifying the gender of a character in a picture, or the like.

[38] The above data label may be understood as a processing result obtained by the target data processing model by processing the target data. For example, it may be result data (corresponding to data of column y in the data processing model) predicted or identified by the data processing model. Specifically, for example, it may be the result "male" or "female" obtained by identifying the gender of a character in a picture through the data processing model.

[39] The above data feature may be understood as a data feature used as a processing basis in the process of processing the target data by the target data processing model. For example, it may be a data feature (corresponding to data of column x in the data processing model) extracted by the data processing model from the sample data and configured to obtain a prediction or identification result. Specifically, for example, it may be a degree of thickness of the hair at the position of a human face in the picture, a bulging amplitude at the position of a human neck in the picture, a color value at the position of the human face in the picture, and so on.

[40] The above evaluation metric (also referred to as an evaluation parameter) may be understood as a parameter capable of evaluating the performance, such as data processing precision and processing efficiency, of the target data processing model.

[41] Specifically, for different types of data processing models, corresponding evaluation metrics may be different. For a classification model, the above evaluation metric may include an Area Under roc Curve (AUC) parameter. The metric is an evaluation metric for measuring a classification effect of a model. Generally, a larger corresponding AUC metric value indicates a better classification effect and a higher precision. In addition, for the classification model, the above evaluation metrics may further include: accuracy, average accuracy, logarithmic loss function, precision-recall, confusion matrix, and the like. For a regression model, the above evaluation metrics may include: square root error, Quantiles of Errors, Almost Correct Predictions, and so on. Of course, it should be noted that the evaluation metrics listed above are only to better illustrate the implementation of the present specification. During a specific implementation, other suitable performance parameters may be selected as the above evaluation metrics according to

specific characteristics of data processing to which the target data processing model is applied and specific requirements of the user. The present specification makes no limitation in this regard.

[42] In this scenario example, the user may set the name to "WM" in a name field of the data processing model in the data input interface presented by the user. In a data label field, data label 1 is set to "man" (for male) and data label 2 is set to "woman" (for female) respectively. In a data feature field, data feature 1 is set to a color value of a human face in the picture, data feature 2 is set to a smoothness value of the human face in the picture, and data feature 3 is set to a bulging amplitude at the position of a human neck in the picture. The evaluation metric is set to "AUC" in the evaluation metric field.

[43] In addition, a sample data input interface is further provided on the above data input interface, through which the user may import corresponding sample data. In the scenario example, the above sample data may specifically be picture data including a character.

[44] The user terminal receives a model description parameter and sample data set by the user through the above data input interface, and sends the above data to the platform server through a network.

[45] After receiving the model description parameter and the sample data, the platform server may automatically perform tests and analyses for the user according to the model description parameter and in conjunction with the sample data, and then determine a model that meets user requirements as a base model (also referred to as an initial model) for the user; and then the base model may be trained using the sample data to construct a target data processing model required by the user and capable of identifying the gender of the character in the picture.

[46] During a specific implementation, after receiving the model description parameter, the server may first parse the model description parameter to determine specific data of the parameters, such as a name of the target data processing model, a data label, a data feature, and an evaluation metric set by the user. For example, it may be determined by parsing that the name of the target data processing model set by the user is "WM"; there are two data labels, and data label 1 is "man"

and data feature 2 is "woman"; there are three data features, wherein data feature 1 is a color value of a human face in the picture, data feature 2 is a smoothness value of the human face in the picture, and data feature 3 is a bulging amplitude at the position of a human neck in the picture; and there is an evaluation metric being an AUC parameter.

[47] Furthermore, by analyzing the above data labels, the server may find that the number of the data labels is 2 instead of 1, and the target data processing model needs to determine whether the target data corresponds to "man" or "woman." Therefore, it can be determined that the data processing to be performed by the target data processing model is data processing of a classification type. Then, preset model types may be searched to determine that a model matching the data processing of the above classification type is the classification type. Therefore, the model type of the base model of the target data processing model is determined as the classification type.

[48] After the model type of the base model is determined as the classification type, the server may search a preset model database, and select a plurality of models by filtering models that meet the model type of the base model (that is, the classification model) to serve as candidate models, so that a model with better effects can be subsequently determined from the above plurality of candidate models to serve as the base model.

[49] In this scenario example, the acquired plurality of models that meet the model type of the base model may be further filtered, so as to select, as far as possible, models with higher probability and relatively good effects by filtering a number of models that meet the model type of the base model to serve as candidate models for a subsequent test analysis, exclude interference models, and avoid a waste of time and resources for testing and analyzing some models with poor effects.

[50] Specifically, in order to more accurately and quickly find better and more suitable candidate models, referring to FIG. 3, the sample data may be used as an input to be input into trained data processing models (for example, data processing model AC, data processing model YU, and data processing model B2) in a preset data processing model library, where their applied base models meet the model type of the above base model, to obtain data processing results (for example,

prediction results) of the data processing models; evaluation metrics of the data processing models are obtained by conducting statistics according to the data processing results of the data processing models; then the data processing model AC whose evaluation metric is similar to or identical with that of the target data processing model is selected as a recommended model according to the evaluation metric of the target data processing model; and base model a_c to which the recommended model AC is applied is determined as a candidate model.

[51] In addition, a plurality of models that meet the model type of the base model may be further filtered according to a data type of the sample data. Referring to FIG. 4, the data type of the sample data may be first determined as picture data. Furthermore, it may be determined that the target data to be processed subsequently by the target data processing model is also picture data. As for models, among different types of models, some have higher accuracy in analyzing and processing picture data, and some have higher accuracy in analyzing and processing text data. In order to improve the accuracy of a data processing model constructed subsequently, models that match the picture data (or are suitable for processing the picture data) may be selected by filtering the plurality of models that meet the model type of the base model to serve as candidate models. For example, models a_c, a_r, b_2, m, and the like may be extracted as candidate models.

[52] Of course, it should be noted that the implementation listed above of further filtering the plurality of models that meet the model type of the base model to obtain a model with a better effect as a candidate model is only a schematic description and should not constitute any improper limitation to the present specification.

[53] After determining a plurality of candidate models, the server may establish a pre-trained model for each of the plurality of candidate models, and then use the above pre-trained model for a test analysis to select a model having a better effect, suitable for target data processing, and meeting user requirements to serve as the base model of the target data processing model.

[54] Specifically, the server may first extract a predetermined proportion of the sample data from the acquired sample data to serve as test data; then pre-process the test data to obtain

processed test data; select, based on the data feature, feature data from the processed test data; and train each of the plurality of candidate models respectively according to the feature data to obtain the pre-trained model.

[55] In this scenario example, during a specific implementation, the server may randomly extract 10% of the sample data to serve as test data. Of course, the value may be set according to specific conditions. For example, if the accuracy of the determined base model is intended to be improved, the proportion value may be increased to increase the amount of the test data. If the determination of the base model is intended to be sped up, the proportion value may be decreased to reduce the amount of the test data. The present specification makes no limitation in this regard.

[56] The above pre-processing may specifically include dimension reduction processing, cleaning processing, and the like. Of course, the pre-processing manners listed above are only to better illustrate the implementation of the present specification. During a specific implementation, a suitable pre-processing manner may be selected according to specific application scenarios and requirements. The present specification makes no limitation in this regard.

[57] In this scenario example, during a specific implementation, the server may specifically extract, in a targeted manner, the part of the data representing the color value of the human face in the picture of the test data as first feature data according to data feature 1, the color value of the human face in the picture. Similarly, the part of the data in the test data that represents the smoothness value of the human face and the bulging amplitude at the position of the human neck may be extracted respectively as second feature data and third feature data. Furthermore, the plurality of candidate models may be trained according to the first feature data, the second feature data, and the third feature data respectively, so as to obtain a plurality of relatively simple pre-trained models. Each pre-trained model is obtained based on a candidate model.

[58] After the plurality of pre-trained models are obtained, the server may further determine, according to the pre-trained models, model parameters corresponding to the pre-trained models through a model parameter search. The above model parameters may be understood as related

operating parameters to be involved in the operation of the pre-trained models. Specifically, the above model parameters may include: the number of times of iterations, a convergence condition, a running time, and so on.

[59] Further, the server may test the pre-trained models. Specifically, adaptabilities of the pre-trained models and the model parameters may be determined according to the evaluation metrics; and then the base model is determined from the plurality of candidate models according to the adaptabilities of the pre-trained models. Therefore, a candidate model suitable for a target data processing scenario, with high accuracy, or meeting user requirements may be selected from the plurality of candidate models to serve as a base model.

[60] The adaptability may be specifically understood as being configured to represent a degree of applicability of a pre-trained model and a model parameter in the data processing on the sample data. In general, for a pre-trained model, a higher adaptability value means that the pre-trained model achieves a higher accuracy and a better effect in performing data processing on the target data. Correspondingly, the candidate model corresponding to the pre-trained model is also more effective in the data processing on the target data.

[61] In this scenario example, during a specific implementation, part of the test data may be randomly extracted to test the established pre-trained models to obtain test results. Statistics is conducted on the AUC parameters of the pre-trained models according to the test results, and the adaptability values of the pre-trained models are determined according to the AUC parameters. For example, if the AUC parameter of a certain pre-trained model is relatively high, the precision of the pre-trained model is relatively high, the applicability in the data processing on the target data is also relatively good, and the adaptability correspondingly obtained is also relatively very high. Then, the plurality of pre-trained models are sorted according to the adaptabilities, and a candidate model corresponding to the pre-trained model with the highest adaptability value (that is, a suitable model) is selected as the base model. Of course, a plurality of candidate models corresponding to a preset number of pre-trained models with top adaptability values or candidate models

corresponding to pre-trained models whose adaptability values are greater than a preset adaptability threshold may also be selected to be combined to obtain a combined model, and the combined model is used as the base model.

[62] For example, pre-trained models 1, 2, 3, 4, 5, and 6 may be used to respectively perform classification identification on the plurality of same pieces of randomly extracted picture data to obtain identification results of the pre-trained models for the picture data, which are used as test results of the pre-trained models. Statistics is then conducted on the AUC parameters of the pre-trained models according to the above test results. The adaptability values of the pre-trained models are further determined according to the AUC parameters. By comparing the adaptability values of the pre-trained models, it is found that the adaptability value of pre-trained model 1 is the highest. Therefore, the candidate model corresponding to pre-trained model 1 may be determined as the base model of the target data processing model.

[63] In this scenario example, the adaptability of each pre-trained model may be further determined in the following manner to evaluate an effect of the corresponding candidate model: performing a parameter search and a structure search according to the pre-trained model and the model parameter to obtain a search result; and determining an adaptability of the pre-trained model and the model parameter according to the search result and the evaluation metric. During a specific implementation, the performing a parameter search and a structure search according to the pre-trained model and the corresponding model parameter may include: performing the parameter search and the structure search through at least one of the following algorithms: a Monte Carlo algorithm, a grid algorithm, a Bayesian algorithm, a genetic algorithm, and the like.

[64] Of course, it should be noted that the method for determining an adaptability of a pre-trained model listed above is only to better illustrate the implementation of the present specification. During a specific implementation, the adaptability of the pre-trained model may be determined by selecting a suitable manner according to a specific situation. The present specification makes no limitation in this regard.

[65] In the above manner, the server may accurately match the model description parameters simply set by the user to a base model suitable for the target data processing model to be constructed by the user, and feed the base model back to the user. Furthermore, the base model may be trained using the sample data to construct a WM model that meets user requirements and can efficiently and accurately identify the gender of the character in the picture, thereby efficiently completing the construction of the target data processing model.

[66] It may be seen from the above scenario example that, according to the method for constructing a data processing model provided in the present specification, by acquiring a model description parameter in a targeted manner, a server can accurately determine a modeling requirement of a user based on the model description parameter, automatically match the modeling requirement of the user to a suitable model as a base model, and then construct a corresponding data processing model, so that it is unnecessary for the user to test and analyze tediously to determine a suitable base model, thereby reducing the operational difficulty of the user and improving the processing efficiency.

[67] Referring to FIG. 5, a method for constructing a data processing model is provided in an embodiment of the present specification, wherein the method is specifically applied to a server side. during a specific implementation, the method may include the following content:

[68] In S51, a model description parameter and sample data of a target data processing model are acquired.

[69] In this implementation, the above target data processing model may specifically be a machine learning model to be constructed for performing target data processing. For example, the above target data processing model may be a semantic recognition model for recognizing text data, a classification model for distinguishing different colors in picture data, or the like. The type and application of the target data processing model are not limited in the present specification.

[70] In this implementation, the above model description parameter may be specifically parameter data that is set by the user and is capable of effectively representing data processing

requirements and/or user customization requirements of a target data processing model to be constructed.

[71] In an implementation, the above model description parameter may specifically include at least one of the following: a name of the target data processing model, a data label, a data feature, and an evaluation metric. Of course, it should be noted that the model description parameters listed above are only to better illustrate the implementation of the present specification. during a specific implementation, other types of parameters may be further introduced as the above model description parameter according to specific application scenarios and user requirements. The present specification makes no limitation in this regard.

[72] The name of the target data processing model may be understood as an identification symbol of the data processing model that the user wants to construct, for example, an A model, a model for identifying the gender of a character in a picture, or the like. The above data label may be understood as a processing result obtained by the target data processing model by processing the target data. For example, it may be result data (corresponding to data of column y in the data processing model) predicted or identified by the data processing model. Specifically, for example, it may be the result "male" or "female" obtained by identifying the gender of a character in a picture through the data processing model. The above data feature may be understood as a data feature used as a processing basis in the process of processing the target data by the target data processing model. For example, it may be a data feature extracted by the data processing model from the sample data and configured to obtain a prediction or identification result (corresponding to data of column x in the data processing model). Specifically, for example, it may be a degree of thickness of the hair at the position of a human face in the picture, a bulging amplitude at the position of a human neck in the picture, a color value at the position of the human face in the picture, and so on. The above evaluation metric (also referred to as an evaluation parameter) may be understood as a parameter capable of evaluating the performance, such as data processing precision and processing efficiency, of the target data processing model. For example, it may be an AUC parameter,

accuracy, average accuracy, logarithmic loss function, precision-recall, confusion matrix, square root error, Quantiles of Errors, Almost Correct Predictions, and the like.

[73] In this implementation, the above sample data may be specifically understood as data of the same type as the to-be-processed target data. The above sample data may specifically be picture data, text data, or the like. The data type of the sample data is not limited in the present specification.

[74] In S53, a base model is determined according to the model description parameter and the sample data.

[75] In this implementation, the above base model may be specifically referred to as an initial model for constructing the target data processing model. The specific base model above may be a model structure (or an algorithm structure) with certain data processing and analysis capabilities. A target data processing model that meets user requirements and is capable of processing target data is generally constructed through training and learning of the base model.

[76] In this implementation, it should be noted that different model structures differ greatly when performing specific data processing. For example, some model structures are suitable for processing text data, while some model structures are suitable for processing picture data. Alternatively, some model structures are suitable for solving classification problems, while some model structures are suitable for solving regression problems. Therefore, determination of a base model that is suitable for target data processing and meets user requirements has a relatively important influence on a use effect of the target data processing model to be subsequently constructed.

[77] In this implementation, in order to avoid that the user tediously performs tests and analyses many times to find a suitable model to serve as the base model, the server may use the model description parameter set by the user as a processing basis, and automatically perform test analyses in conjunction with the sample data input by the user to search for a model (that is, a suitable

model) that can perform target data processing and meet user requirements to serve as a data processing model.

[78] In this implementation, during a specific implementation, a model feature of the target data processing model to be constructed by the user may be determined according to the model description parameter; and then according to the model feature, the range is gradually narrowed down to filter a plurality of model structures to obtain a suitable model as a base model, which is provided for the user to construct a specific target data processing model.

[79] In an implementation, during a specific implementation, the determining a base model according to the model description parameter and the sample data includes: determining a model type of the base model from a preset model type group according to the data label, wherein the preset model type group includes at least a classification type and a regression prediction type; selecting a plurality of models from models meeting the model type of the base model to serve as candidate models; establishing a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature, and determining a model parameter of each pre-trained model; determining an adaptability of each pre-trained model according to the evaluation metric, wherein the adaptability is configured to represent a degree of applicability of data processing on the sample data based on a pre-trained model and a model parameter corresponding to the pre-trained model; and determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models.

[80] In S55, the base model is trained according to the sample data to obtain the target data processing model.

[81] In this implementation, after the base model is determined, the base model may be further used as an initial model; and then the base model is trained using the sample data to construct a target data processing model that meets requirements. Of course, it should be noted that the constructing a target data processing model by training using the base model listed above is only an implementation of applying the base model. During a specific implementation, the determined

base model may also be used for other applications according to a specific application scenario and a degree of technological development. The present specification makes no limitation in this regard.

[82] In this implementation, by separating the model description and the determination of the base model from the construction of the data processing model, the user only needs to provide a model description parameter, and the server can automatically perform the model description according to the model description parameter provided by the user and automatically determine a suitable base model which is provided for the user, so that the data processing model can be constructed based on the above base model. Therefore, the user does not need to tediously search for a suitable base model by himself/herself, thereby reducing the burden of the user.

[83] It may be seen from the above that, according to the method for constructing a data processing model provided in the present specification, by acquiring a model description parameter in a targeted manner, a server can accurately determine a modeling requirement of a user based on the model description parameter, automatically match the modeling requirement of the user to a suitable model as a base model, and then construct a corresponding data processing model, so that it is unnecessary for the user to test and analyze tediously to determine a suitable base model, thereby reducing the operational difficulty of the user and improving the processing efficiency.

[84] In an implementation, the model description parameter may specifically include at least one of the following: a name of the target data processing model, a data label, a data feature, an evaluation metric, and the like. Of course, it should be noted that the model description parameter listed above is only a schematic illustration. During a specific implementation, other types of parameters may be further introduced as the above model description parameter according to a specific situation. The present specification makes no limitation in this regard.

[85] In an implementation, the model description parameter may further specifically include: a model type of the base model, a split rule of the sample data, a threshold parameter, and the like. For users with certain professional capabilities, such as engineers, setting permissions of more

model parameters may be provided for them to acquire more model parameters, so that a suitable model can be more quickly and accurately found based on more model parameters to serve as the base model of the target data processing model.

[86] In an implementation, during a specific implementation, determining a base model according to the model description parameter and the sample data may include the following content:

[87] In S1, a model type of the base model is determined from a preset model type group according to the data label, wherein the preset model type group includes at least a classification type and a regression prediction type.

[88] In this implementation, the determining a model type of the base model from a preset model type group according to the data label class may be specifically understood as analyzing the number of the data labels and a relationship between data of the data label and the data label to determine a type of target data processing to be performed by the target data processing model to be constructed; and then determining the type of the base model applicable to the target data processing according to the type of the target data processing.

[89] In this implementation, it should be further added that the model type may also be determined by analyzing the evaluation metric. Therefore, during a specific implementation, in addition to the data label, a model type of the base model may also be determined from the preset model type group according to the evaluation metric, or a combination of the data label and the evaluation metric.

[90] In this implementation, the above preset model type group includes at least a classification type, a regression prediction type, and the like. Each type may include a plurality of specific model structures. Of course, it should be noted that the model types included in the preset model type group listed above are only a schematic illustration. According to a specific application scenario and the development of technology, the above preset model type group may also include other model types. The present specification makes no limitation in this regard.

[91] In this embodiment, it should be added that a model whose model type is a classification type is more suitable for data processing involving classification problems, and a model whose classification type is a regression prediction model is often more suitable for data processing involving prediction problems.

[92] In S2, a plurality of models are selected from models meeting the model type of the base model to serve as candidate models.

[93] In this implementation, the above candidate models may be specifically understood as model structures that are suitable for target data processing (that is, meeting the model type of the base model) and may be used as the base model of the target data processing model. Specifically, the candidate models may be further tested and filtered subsequently to determine the base model from the candidate models.

[94] In S3, a pre-trained model is established for each of the plurality of candidate models according to the sample data and the data feature, and a model parameter of each pre-trained model is determined.

[95] In this implementation, the above pre-trained model may be specifically understood as a relatively simple data processing model obtained based on the candidate models.

[96] In this implementation, the above model parameter may be specifically understood as related operating parameters to be involved in the operation of the pre-trained models. Specifically, the above model parameters may include: the number of times of iterations, a convergence condition, a running time, and so on. Of course, the model parameters listed above are only a schematic illustration. The specific content of the model parameters is not limited in the present specification.

[97] In this embodiment, the establishing a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature may specifically include: extracting test data from the sample data; extracting feature data from the test data according to the

data feature; and training each candidate model by using the feature data to obtain a pre-trained model corresponding to each candidate model.

[98] The above feature data may be understood as partial data that includes the data feature in the sample data. For example, the above feature data may be data that represents the color of the human face in the picture.

[99] In this implementation, after the plurality of pre-trained models are obtained, model parameters corresponding to the pre-trained models may further be determined through a model parameter search according to the specific pre-trained models.

[100] In S4, an adaptability of each pre-trained model is determined according to the evaluation metric, wherein the adaptability is configured to represent a degree of applicability of data processing on the sample data based on a pre-trained model and a model parameter corresponding to the pre-trained model.

[101] In this implementation, the adaptability may be specifically understood as being configured to represent a degree of applicability of a pre-trained model and a model parameter in the data processing on the sample data. In general, for a pre-trained model and a corresponding model parameter, the higher the adaptability value is, the higher the accuracy of the pre-trained model in the data processing on the target data based on the corresponding model parameter is, and the better the effect is. Correspondingly, the candidate model corresponding to the pre-trained model is also more effective in the data processing on the target data based on the above model parameter.

[102] In an implementation, the determining an adaptability of each pre-trained model according to the evaluation metric may specifically include the following content: processing test data by using the pre-trained model and the corresponding model parameter respectively to obtain a data processing result to serve as a test result of each pre-trained model and the model parameter; conducting statistics on the evaluation metric of each pre-trained model according to the test result of each pre-trained model and the model parameter; and determining an adaptability of each pre-trained model and the model parameter according to a metric parameter of each pre-trained model.

It should be noted that the manner of determining an adaptability listed above is only a schematic illustration. During a specific implementation, the adaptability of the pre-trained model may also be determined in other suitable manners according to specific situations and requirements. The present specification makes no limitation in this regard.

[103] In S5, the base model is determined from the plurality of candidate models according to the adaptabilities of the pre-trained models.

[104] In this implementation, during a specific implementation, the determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models may include: comparing the adaptabilities of the pre-trained models to determine a model corresponding to the pre-trained model with the highest adaptability as the base model of the target data processing model. It may also include: combining models corresponding to the plurality of pre-trained models whose adaptabilities are greater than a preset adaptability threshold to obtain a combined model; determining the combined model as the base model, and so on. Of course, the determination of a suitable base model from the candidate models according to the adaptabilities listed above is only a schematic illustration. During a specific implementation, the base model may also be determined according to the adaptabilities in other suitable manners depending on a specific situation. The present specification makes no limitation in this regard.

[105] In an implementation, during a specific implementation, selecting a plurality of models from models meeting the model type of the base model to serve as candidate models may include the following content:

[106] In S1, data processing is performed on the sample data by using data processing models in which applied base models meet the model type of the base model in a preset data processing model library to obtain data processing results of the data processing models.

[107] In S2, statistics is conducted on evaluation metrics of the data processing models according to the data processing results of the data processing models.

[108] In S3, a data processing model in which a degree of difference between its evaluation metric and the evaluation metric of the target data processing model is less than a preset degree threshold is determined as a recommended model.

[109] In S4, a base model to which the recommended model is applied is determined as a candidate model.

[110] In this implementation, a variety of different trained data processing models are specifically stored in the preset data processing model library. It should be added that the above preset data processing model library may acquire the latest trained data processing model after every preset time period to update the preset data processing model library.

[111] In this implementation, a degree of difference between the above evaluation metric and the evaluation metric of the target data processing model being less than a preset degree threshold may be specifically understood as that the evaluation metric of the data processing model is similar to or identical with the evaluation metric of the target data processing model. The above preset degree threshold may be set flexibly according to the precision requirements. The present specification makes no limitation in this regard.

[112] In this implementation, the above recommended model may be understood as a trained data processing model of which applicable data processing is similar to or identical with the target data processing to be performed by the target data processing model to be constructed. Such a data processing model has a high reference value for constructing the target data processing model.

[113] In this implementation, during a specific implementation, the acquired sample data may be used as an input to be input into data processing models in a preset data processing model, in which the applied base models meet the model type of the above base model, to obtain data processing results of the data processing models; evaluation metrics of the data processing models are obtained by conducting statistics according to the data processing results of the data processing models respectively; then a data processing model whose evaluation metric is identical with or similar to that of the target data processing model is found and determined as a recommended model by

taking the evaluation metric of the target data processing model as a basis; and the model to which the recommended model is applied is determined as a candidate model, so that the candidate model with a better effect can be found more quickly and efficiently.

[114] In an implementation, in order to further reduce the number of the candidate models and improve the efficiency of determining the base model, during a specific implementation, the method may further include the following content:

[115] In S1, a data type of the sample data is determined, wherein the data type of the sample data includes at least one of the following: picture data and text data.

[116] In S2, a model type matching the data type of the sample data is filtered from the model types of the base models according to the data type of the sample data to serve as the model type of the base model.

[117] In this implementation, the data type of the sample data may specifically include picture data, text data, image data, numeric data, and so on. Of course, it should be noted that the data types listed above are only to better illustrate the implementation of the present specification. The data type of the sample data is not limited in the present specification.

[118] In this implementation, it should be added that different model structures are often applicable to different data types. For example, some models may be suitable for processing image data, but not suitable for processing text data. Some models may be suitable for processing text data, but not suitable for processing numeric data. Therefore, in order to further reduce the number of the candidate models and more quickly determine a base model suitable for the target data processing, model types of base models may be further limited in conjunction with the data type of the sample data. For example, if a problem involved in data processing is a classification problem and the data type of the sample data is text data, the model type of the base model may be determined, by taking the above into account, as a classification type and a model type suitable for processing text data (that is, matching text data). In this way, the range of types of models that may be used as a base model may be further narrowed, so that the number of the candidate models

acquired based on the model type of the base model can be reduced, and a suitable base model can be determined only by testing fewer candidate models subsequently. Thus, the objective of reducing the workload and improving the determination efficiency is achieved.

[119] In an implementation, during a specific implementation, the establishing a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature may include the following content:

[120] In S1, a predetermined proportion of sample data is extracted from the sample data to serve as test data.

[121] In this implementation, during a specific implementation, part of the sample data may be randomly extracted from the sample data according to a preset proportion to serve as the test data. The above preset proportion may be set flexibly according to a specific application scenario and precision requirements. For example, if the accuracy of the determined base model is intended to be improved, the proportion value may be increased to increase the amount of the test data. If the determination of the base model is intended to be sped up, the proportion value may be decreased to reduce the amount of the test data. The present specification makes no limitation in this regard.

[122] In S2, the test data is pre-processed to obtain processed test data, wherein the pre-processing includes: dimension reduction processing and/or cleaning processing.

[123] In this implementation, the above pre-processing may specifically include dimension reduction processing, cleaning processing, and the like. Of course, the pre-processing manners listed above are only to better illustrate the implementation of the present specification. During a specific implementation, a suitable pre-processing manner may be selected according to specific application scenarios and requirements. For example, the sample data has a large amount of data and includes much information. Therefore, the test data may be pre-processed through dimension reduction processing to improve the efficiency of subsequent data processing. For another example, for some data processing with strong randomness, it is required that there should be no sequential connection between the sample data for training as far as possible. Therefore, the test

data may be pre-processed by cleaning the data first to reduce a model error caused by an interrelation between the test data.

[124] In S3, feature data is selected from the processed test data according to the data feature.

[125] In this implementation, selecting feature data from the processed test data according to the data feature may specifically include: extracting partial data including the data feature from the processed test data to serve as the feature data (which may also be referred to as feature engineering).

[126] In S4, each of the plurality of candidate models is trained according to the feature data respectively to obtain the pre-trained model.

[127] In this implementation, the above pre-trained model may be specifically understood as a data processing model obtained through a relatively simple training based on a candidate model and having certain data processing precision.

[128] In this implementation, training each of the plurality of candidate models according to the feature data respectively to obtain the pre-trained model may specifically include: training the candidate models respectively by using the extracted feature data, so as to obtain a plurality of relatively simple pre-trained models. The above pre-trained models each correspond to a candidate model.

[129] In an implementation, during a specific implementation, the determining an adaptability of each pre-trained model according to the evaluation metric may include the following content: performing a parameter search and a structure search according to the pre-trained model and the model parameter to obtain a search result; and determining an adaptability of the pre-trained model and the model parameter according to the search result and the evaluation metric.

[130] In this implementation, the performing a parameter search and a structure search according to the pre-trained model and the corresponding model parameter may specifically include: performing the parameter search and the structure search through at least one of the following algorithms: a Monte Carlo algorithm, a grid algorithm, a Bayesian algorithm, a genetic algorithm,

and the like. Of course, it should be noted that the algorithms listed above for performing the parameter search and the structure search are only a schematic illustration. During a specific implementation, the parameter search and the structure search may also be performed according to other algorithms depending on a specific application scenario and precision requirements. The present specification makes no limitation in this regard.

[131] In this implementation, during a specific implementation, the determining the adaptability of the pre-trained model and the model parameter according to the search result and the evaluation metric may include: scoring each pre-trained model and the corresponding model parameter according to the search result of each pre-trained model and the model parameter as well as an evaluation metric set by the user, and taking the score value as the adaptability of the pre-trained model and the model parameter.

[132] In an implementation, during a specific implementation, the determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models may include: determining a model used by the pre-trained model with the highest adaptability in the pre-trained models as the base model.

[133] In an implementation, during a specific implementation, the determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models may further include: combining models used by a plurality of pre-trained models, whose adaptabilities are greater than a preset adaptability threshold (or whose adaptabilities are ranked at the top), in the pre-trained models to obtain a combined model; and determining the combined model as the base model. In this way, the advantages of a variety of different model structures with better effects can be synthesized to obtain a model structure that is more suitable for the target data processing to serve as a base model.

[134] In an implementation, the model description parameter may further include: a model type of the base model, a split rule of the sample data, and the like. The above split rule of the sample data may specifically include a preset proportion of the test data set by the user, a specified manner

for splitting the sample data, and the like. In this way, more optional model description parameters can be provided for users with different professional capabilities for setting, so that personalized requirements of the users can be better met, and a base model and a model parameter can be determined more accurately and quickly.

[135] It may be seen from the above that, according to the method for constructing a data processing model provided in the present specification, by acquiring a model description parameter in a targeted manner, a server can accurately determine a modeling requirement of a user based on the model description parameter, automatically match the modeling requirement of the user to a suitable model as a base model, and then construct a corresponding data processing model, so that it is unnecessary for the user to test and analyze tediously to determine a suitable base model, thereby reducing the operational difficulty of the user and improving the processing efficiency. In addition, the base model is determined by selecting a suitable candidate model according to a variety of features, such as the data label and the data type of the sample data, which eliminates some interference and reduces the workload, thereby further improving the processing efficiency.

[136] Referring to FIG. 6, a data processing method for performing specific target data processing is provided in an embodiment of the present specification. During a specific implementation, the method may include the following content:

[137] In S61, to-be-processed target data is acquired.

[138] In this implementation, the to-be-processed target data may specifically be picture data to be classified, text data to be identified, image data to be predicted, or the like. The data type and data content of the to-be-processed target data are not limited in the present specification.

[139] In S63, a target data processing model is invoked to process the target data, wherein the target data processing model is obtained by determining a base model according to a model description parameter and sample data of the target data processing model that are acquired and then training the base model according to the sample data.

[140] In this implementation, the above target data processing model is a data processing model constructed in a manner that a server obtains, according to a model description parameter set and sample data input by a user in advance and through matching, a suitable model suitable for target data processing and meeting user requirements to serve as a base model; and then trains and learns the base model by using the sample data.

[141] In this implementation, the invoking a target data processing model to process the target data may specifically include: taking the target data as an input, and inputting the target data into the target data processing model; and performing, by the target data processing model, the corresponding data processing of the target data processing model to obtain a corresponding data processing result. For example, a picture including a character is input into a target data processing model capable of identifying the gender of the character in the picture; and the target data processing model performs corresponding data processing on the picture to obtain a data processing result as follows: the character in the picture is a female, thereby completing data processing on target data.

[142] Referring to FIG. 7, a data processing method is provided in an embodiment of the present specification. The method may be applied to a user terminal to acquire a model description parameter of a target data processing model. During a specific implementation, the method may include the following content:

[143] In S71, a data input interface is displayed.

[144] In this implementation, the above data input interface may specifically be an image input interface as shown in FIG. 2, a code input interface (that is, a code template provided for the user to set definitions), or the like. Of course, it should be noted that the data input interface listed above is only to better illustrate the implementation of the present specification. The specific form of the data input interface is not limited in the present specification.

[145] In this implementation, it should be added that inputting a model description parameter through an input interface is only a schematic illustration. During a specific implementation, the

user may also input a model parameter of the target data processing model through a command-line interaction such as presto (a distributed SQL query engine) and ipython-notebook (a Python-based programming tool), or an input interface incorporated into an API or a program.

[146] In S73, a model description parameter and sample data of a target data processing model input by a user on the data input interface are received, wherein the model description parameter and the sample data are configured to determine a base model, and the base model is configured to establish the target data processing model.

[147] In this implementation, it should be noted that, in existing methods for constructing a data processing model, model description, determination of a base model, and model training and construction are mixed together, and as a result, during the implementation, the methods are inflexible, are not easy to expand, are not friendly enough to developers, are incapable of tracking the entire development process, cannot perform flexible help control and code review by means of a version management tool, such as git, and are not easy to update a new solution model for a certain problem. That is, the existing methods are relatively difficult for the user to operate, and are not convenient and friendly enough.

[148] In order to solve the above problems, a model description parameter may be defined in advance through language definition, so that a user can input or set, in a targeted manner, a specific model description parameter that meets target data processing and user customization requirements. A machine may follow a certain syntax rule to acquire the specific model description parameter set by the user and generate a corresponding execution file, such as a code program, and then automatically find a suitable base model according to the model description parameter set by the user in the execution file and provide it to the user; and then train the found base model to construct a specific target data processing model, thereby separating the model description and the determination of the base model from the model training and construction, reducing the difficulty in and complexity of constructing the data processing model, and improving the user experience.

[149] In this implementation, during a specific implementation, the above model description parameter may be defined in a Structured Query Language (SQL) according to a certain syntax rule (for example, by way of keywords). For example, code for definition about a model description parameter is generated, and then a corresponding data input interface is displayed to the user based on the above code for definition, so as to receive the model description parameter set by the user. Of course, it should be noted that the definition language and the syntax rule listed above are only to better illustrate the implementation of the present specification. during a specific implementation, the above model description parameter may also be defined in other languages or syntax rules according to a specific situation. The present specification makes no limitation in this regard.

[150] In this implementation, the above model description parameter may specifically include: a name of the target data processing model, a data label, a data feature, an evaluation metric, and the like, and corresponding keywords are model_name, label, feature, eval_metrics, and the like respectively. Of course, it should be noted that the model description parameters listed above are only to better illustrate the implementation of the present specification. They should not constitute any improper limitation to the present specification.

[151] In this implementation, during a specific implementation, a specific definition may be made in the following manner to obtain code for a definition including a model description parameter: TRAIN model_name [[BY model_type] [WITH params]] [FOLD number] [PARTITION partition_fields]
SELECT [LABEL label_names]
EVAL eval_metrics [SELECT]
[TIME time_limit].

[152] The above code for definition includes four parts, that is, a training part, a data part, an evaluation part, and a computation limiting condition part.

[153] The above TRAIN model_name [[BY model_type] [WITH params]] [FOLD number] [PARTITION partition_fields]
 is used to define the training part, where [] represents an optional part, TRAIN represents a keyword of the training part, and model_name represents a

model name generated by training; BY is used to specify a model type model_type for training; WITH is used to specify model parameters params; FOLD is used to specify a value number of the layer k-fold; and PARTITION is used to specify data field partition_fields for partitioning. It should be added that the specific meaning of the above
 can be obtained with reference to the related art, such as HTML 5. This may be expressed as a common description, such as dividing a problem into a plurality of sub-problems and dividing a model into sub-models.

[154] The above SELECT [LABEL label_names]
 is used to define the data part, where [] represents an optional part, SELECT represents the use of one or more data fields, and LABEL is used to specify one or more labels label_names. For example, a first data field may be used as a label LABEL by default. If there is more than one LABEL, label_names
 may be specified using LABEL.

[155] The above EVAL eval_metrics [SELECT]
 is used to define the evaluation part, where [] represents an optional part, and EVAL is used to specify an evaluation metric eval_metrics. If different data is used for evaluation, an evaluation data part
 may be further added to the evaluation part.

[156] The above [TIME time_limit] is used to define the computation limiting condition part, where [] represents an optional part, and TIME is used to specify a time limit for the completion of model training. For example, TIME may be used to specify a time limit for the completion of model training; and if the time limit is not specified, it can be completely processed by the system.

[157] Further, in consideration of offline computing for the training model, the above code for definition may also include an offline computing part, which is expressed as:
SCORE mode_name TO table_name INDEX Index_list (create a table/insert with index list)
 SELECT
.
SCORE is used to select a trained model name; mode_name TO is used to define a data table table_name stored after the computing; INDEX is used to define an index field Index_list of the data table, and the index fields may also appear in the SELECT part; and
SELECT is used to define data for offline computing. For example, it may be required that the non-LABEL field of the training model be exactly the same after removal of the INDEX part.

[158] Further, in consideration of online computing for the training model, the above code for definition may also include an online computing part, which is expressed as:
DEPLOY model_name WITH params
. DEPLOY is used to deploy the training model model_name as an interface service (for example, rest api), and WITH is used to add a corresponding parameter configuration params according to requirements of a data center. The
 language may be expanded by nesting to a situation where a plurality of models are combined.

[159] Of course, the code for definition listed above is only a schematic illustration, and should not constitute any improper limitation to the present specification. The specific description of the language may be evolving or changing. For example, the following forms may also be described:

CREATE TABLE model1 AS

SELECT

model_train (features [, labels] [, options]) FROM source_table1;

CREATE TABLE predict1 ASSELECT model_predict (model1 [, options]) FROM source_table2.

[160] In this implementation, for the image input interface, as shown in FIG. 2, the model description parameter and the sample data of the target data processing model may be acquired by receiving data information in each data field. For example, a data label set by the user may be acquired through a data label field in the data input interface, a data feature set by the user may be acquired through a data feature field, an evaluation metric set by the user may be acquired through an evaluation metric field, a model name of the target data processing model may be acquired through a name field of the target data processing model, and sample data input by the user may be acquired through a sample data input interface.

[161] In this implementation, for the code input interface, the model description parameter of the target data processing model may be acquired by receiving a character string set by the user at a

preset position in the code input interface, and the corresponding sample data at a specified position is acquired by receiving a preset code instruction triggered by the user.

[162] Specifically, the user may define or set a specific character string at a preset position in the above code input interface to serve as the model description parameter. For example, the user terminal may receive, through the SELECT instruction, a character string set by the user in the underlined position of "TRAIN model_name" in the code input interface to serve as the model name of the target data processing model; receive, through the SELECT instruction, a character string set by the user in the underlined position of "LABEL label_names" to serve as the data label of the target data processing model; and receive, through the SELECT instruction, a character string set by the user in the underlined position of "EVAL eval_metrics" to serve as the evaluation metric of the target data processing model. The acquisition of other model description parameters may be obtained with reference to the above implementation, and the descriptions thereof are omitted in the present specification.

[163] In this implementation, after acquiring the model description parameter and the sample data in the above manner, the user terminal may send the above model description parameter and the sample data to the server in the form of an execution file such as code. The server may acquire the model description parameter set by the user by parsing (SQL-based parsing) the execution file such as code; determine a base model according to the model description parameter set by the user; and then train the base model according to the sample data to obtain the target data processing model.

[164] In this implementation, when the server receives an execution file such as code including model description parameters, during a specific implementation, the server may correspondingly divide the execution file into four parts for analysis and processing.

[165] The first part is a training part. Specifically, a keyword TRAIN may be used to extract model_name to serve as a model name of a data processing model generated by training; which model_types (that is, model types) are used for training may be specified using a keyword BY; parameters params of a related model may be specified using a keyword WITH, wherein the model

parameter may specifically specify the value of k-fold by using FOLD described by Json; and data fields partition_fields for partitioning may be specified using a keyword PARTITION.

[166] The second part is a data part. Specifically, a first data field may be used as a label LABEL by default through a SQL syntax by using a keyword SELECT. Specifically, if there is more than one LABEL, label_names are specified using LABEL.

[167] The third part is an evaluation part. Specifically, an evaluation metric eval_metrics may be specified using a keyword EVAL. If different data is used for evaluation, an evaluation data part may be further added to the evaluation part.

[168] The fourth part is a computation limiting condition part. Specifically, a time limit for the completion of model training may be specified using a keyword TIME. If the user does not set a time limit, it may be completely processed by the server.

[169] Further, for the server, a trained model may also be used for offline computing in the following manner: CORE mode_name TO table_name INDEX Index_list
SELECT
. The offline computing of the model may be specifically divided into two parts: SCORE is used to select a trained model name mode_name; TO is used to define a data table stored after the computing; and INDEX is used to define an index field of the data table, wherein the index fields may also appear in the SELECT part. SELECT is used to define data for offline computing, which is required to be, after removal of the INDEX part, exactly the same as the non-label field of the training model. The online computing of the model may be performed in the following manner: DEPLOY model_name WITH params deploys the model into a rest api service, and WITH adds a corresponding parameter configuration params according to requirements of a data center.

[170] In this way, the server may acquire the specific model description parameter set by the user through parsing, so that a suitable base model can be automatically determined subsequently by using the above model description parameter.

[171] In an implementation, in consideration of some users with relatively strong professional capabilities, they may be provided with more freedom and authority to set model description

parameters, so that the server can more quickly and accurately determine a suitable base model. Therefore, after receiving the model description parameter and the sample data of the target data processing model input by the user on the data input interface, during a specific implementation, the method may further include the following content: displaying, in response to a display operation, an execution file generated based on the model description parameter; and receiving a modification operation on the execution file, and updating the execution file according to the modification operation.

[172] In this implementation, the above execution file may be understood as specific program code. The above execution file includes model description parameters and other model setting parameters automatically generated based on the model description parameters.

[173] In this implementation, the above data input interface may specifically include a display key or an instruction for displaying an execution file. The user may send the display operation by clicking the above display key or invoking the instruction of displaying an execution file. Then, the user terminal displays the corresponding execution file after receiving the display operation from the user.

[174] The user may directly perform a modification operation on the model description parameters and other model setting parameters included in the above execution file. The user terminal may receive a modification operation of a user on an execution file, update the execution file according to the modification operation of the user, and send the updated execution file to a server, so that the server can more quickly and accurately determine a base model according to the execution file modified by the user to construct a target data processing model.

[175] A server is further provided in an embodiment of the present specification, including a processor and a memory for storing processor-executable instructions, wherein, during a specific implementation, the processor may perform the following steps according to the instructions: acquiring a model description parameter and sample data of a target data processing model;

determining a base model according to the model description parameter and the sample data; and training the base model according to the sample data to obtain the target data processing model.

[176] In order to complete the above instructions more accurately, as shown in FIG. 8, another specific server is further provided in the present specification, wherein the server includes network communication port 801, processor 802, and memory 803. The above structures are connected to each other by internal cables, so that the structures can perform specific data interaction.

[177] Network communication port 801 may be specifically configured to acquire a model description parameter and sample data of a target data processing model.

[178] Processor 802 may be specifically configured to determine a base model according to the model description parameter and the sample data; and train the base model according to the sample data to obtain the target data processing model.

[179] Memory 803 may be specifically configured to store the acquired model description parameter and sample data, as well as intermediate data generated by processor 902, and store a corresponding instruction program.

[180] In this implementation, network communication port 801 may be a virtual port that is bound to different communication protocols so that different data can be sent or received. For example, the network communication port may be port 80 responsible for web data communication, port 21 responsible for FTP data communication, or port 25 responsible for mail data communication. In addition, the network communication port may also be a physical communication interface or a communication chip. For example, it may be a wireless mobile network communication chip, such as GSM, CDMA, or the like; or a Wifi chip; or a Bluetooth chip.

[181] In this implementation, processor 802 may be implemented in any suitable manner. For example, the processor may be in the form of, for example, a microprocessor or processor and a computer-readable medium storing computer-readable program code (such as software or firmware) executable by the (micro)processor, a logic gate, a switch, an Application Specific

Integrated Circuit (ASIC), a programmable logic controller, an embedded micro-controller, or the like. The present specification makes no limitation thereto.

[182] In this implementation, memory 803 may include a plurality of levels. In a digital system, any device may be a memory as long as it can store binary data. In an integrated circuit, a circuit that has a storage function but does not have a physical form may also be referred to as a memory, such as a RAM or a FIFO; in the system, a storage device in a physical form may also be referred to as a memory, such as a memory bank and a TF card.

[183] A user terminal is further provided in an embodiment of the present specification, including a processor and a memory for storing processor-executable instructions, wherein, during a specific implementation, the processor may perform the following steps according to the instructions: displaying a data input interface; and receiving a model description parameter and sample data of a target data processing model input by a user on the data input interface, wherein the model description parameter and the sample data are configured to determine a base model, and the base model is configured to establish the target data processing model.

[184] In order to complete the above instructions more accurately, as shown in FIG. 9, another specific user terminal is further provided in the present specification, wherein the user terminal includes display 901, processor 902, and memory 903. The above structures are connected to each other by internal cables, so that the structures can perform specific data interaction.

[185] Display 901 may be specifically configured to display a data input interface.

[186] Processor 902 may be specifically configured to receive a model description parameter and sample data of a target data processing model input by a user on the data input interface, wherein the model description parameter and the sample data are configured to determine a base model, and the base model is configured to establish the target data processing model.

[187] Memory 903 may be specifically configured to store a corresponding instruction program.

[188] In this implementation, display 901 may be a display tool that displays a certain electronic file on a screen through a specific transmission device and then reflects it to human eyes. The

display may include a liquid crystal display (LCD) screen, a cathode ray tube (CRT) display screen, a light-emitting diode (LED) display screen, and the like.

[189] A computer storage medium based on the above method for constructing a data processing model is further provided in an embodiment of the present specification. The computer storage medium stores computer program instructions thereon, wherein, when the computer program instructions are executed, the following operations are implemented: acquiring a model description parameter and sample data of a target data processing model; determining a base model according to the model description parameter and the sample data; and training the base model according to the sample data to obtain the target data processing model.

[190] In this implementation, the above storage medium includes, but is not limited to, a Random Access Memory (RAM), a Read-Only Memory (ROM), a Cache, a Hard Disk Drive (HDD), or a Memory Card. The memory may be configured to store computer program instructions. A network communication unit may be an interface set according to a standard specified by a communication protocol and configured to perform network connection communication.

[191] In this implementation, the functions and effects specifically implemented by the program instructions stored in the computer storage medium can be explained in comparison with other embodiments, and the descriptions thereof are omitted here.

[192] Referring to FIG. 10, at the software level, an apparatus for constructing a data processing model is further provided in an embodiment of the present specification. The apparatus may specifically include the following structural modules:

[193] acquisition module 1001 which may be specifically configured to acquire a model description parameter and sample data of a target data processing model;

[194] determination module 1002 which may be specifically configured to determine a base model according to the model description parameter and the sample data; and

[195] training module 1003 which may be specifically configured to train the base model according to the sample data to obtain the target data processing model.

[196] In an implementation, the model description parameter may specifically include at least one of the following: a name of the target data processing model, a data label, a data feature, an evaluation metric, and the like. Of course, it should be noted that the model description parameter listed above is only a schematic illustration. During a specific implementation, other types of parameters may be further introduced as the above model description parameter. The present specification makes no limitation in this regard.

[197] In an implementation, determination module 1002 may specifically include the following structural units:

[198] a first determination unit which may be specifically configured to determine a model type of the base model from a preset model type group according to the data label, wherein the preset model type group includes at least a classification type and a regression prediction type;

[199] a selection unit which may be specifically configured to select a plurality of models from models meeting the model type of the base model to serve as candidate models;

[200] an establishment unit which may be specifically configured to establish a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature, and determine a model parameter of each pre-trained model;

[201] a second determination unit which may be specifically configured to determine an adaptability of each pre-trained model according to the evaluation metric, wherein the adaptability is configured to represent a degree of applicability of data processing on the sample data based on a pre-trained model and a model parameter corresponding to the pre-trained model; and

[202] a third determination unit which may be specifically configured to determine the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models.

[203] In an implementation, the selection unit may specifically include the following structural sub-units:

[204] a processing sub-unit which may be specifically configured to process the sample data by using data processing models in a preset data processing model library whose base models meet the model type of the base model to obtain data processing results of the data processing models;

[205] a statistics sub-unit which may be specifically configured to conduct statistics on evaluation metrics of the data processing models according to the data processing results of the data processing models;

[206] a first determination sub-unit which may be specifically configured to determine a data processing model in which a degree of difference between its evaluation metric and the evaluation metric of the target data processing model is less than a preset degree threshold as a recommended model; and

[207] a second determination sub-unit which may be specifically configured to determine a base model to which the recommended model is applied as a candidate model.

[208] In an implementation, the establishment unit may specifically include the following structural sub-units:

[209] an extraction sub-unit which may be specifically configured to extract a preset proportion of sample data from the sample data to serve as test data;

[210] a pre-processing sub-unit which may be specifically configured to pre-process the test data to obtain processed test data, wherein the pre-processing includes: dimension reduction processing and/or cleaning processing;

[211] a selection sub-unit which may be specifically configured to select feature data from the processed test data according to the data feature; and

[212] a training sub-unit which may be specifically configured to train each of the plurality of candidate models according to the feature data respectively to obtain the pre-trained model.

[213] In an implementation, the third determination unit may specifically include: a third determination sub-unit which may be specifically configured to determine a model used by the pre-trained model with the highest adaptability as the base model.

[214] In an implementation, the third determination unit may further include: a fourth determination sub-unit which may be specifically configured to combine models used by the plurality of pre-trained models whose adaptabilities are greater than a preset adaptability threshold to obtain a combined model; and determine the combined model as the base model.

[215] In an implementation, the model description parameter may further include: a model type of the base model, a split rule of the sample data, and the like. In this way, more optional model description parameters can be provided for users with different professional capabilities, so that a base model and a model parameter can be more accurately and quickly determined.

[216] It should be noted that the units, apparatuses, or modules illustrated in the above embodiments may be specifically implemented by a computer chip or an entity, or implemented by a product having a certain function. For ease of description, when the above apparatus is described, it is divided into various modules based on functions for separate description. Of course, during an implementation of the present specification, the functions of the modules may be implemented in one or more pieces of software and/or hardware, or the modules that implement the same function may be implemented by a combination of multiple sub-modules or sub-units. The apparatus embodiments described above are only schematic. For example, the division of the units is only a division of logical functions. In actual implementations, there may be other division manners. For example, a plurality of units or components may be combined or integrated into another system, or some features can be ignored or not implemented. In addition, the displayed or discussed mutual coupling or direct coupling or communication connection may be indirect coupling or communication connection through some interfaces, apparatuses, or units, and may be in electrical, mechanical, or other forms.

[217] It may be seen from the above that, according to the apparatus for constructing a data processing model provided in the embodiment of the present specification, an acquisition module acquires a model description parameter in a targeted manner, a determination module accurately determines a modeling requirement of a user based on the model description parameter and

automatically matches the modeling requirement of the user to a suitable model as a base model, and then a training model constructs a corresponding data processing model, so that it is unnecessary for the user to test and analyze tediously to determine a suitable base model, thereby reducing the operational difficulty of the user and improving the processing efficiency.

[218] Although the present specification provides method operation steps as described in the embodiments or flowcharts, more or fewer operation steps may be included based on conventional or non-creative means. The sequence of steps listed in the embodiments is only one of many sequences of execution of the steps, which does not represent a unique sequence of execution. When executed by an actual apparatus or client terminal product, the steps may be executed sequentially according to the method shown in the embodiments or the drawings or in parallel (for example, a parallel processor or multi-thread processing environment, or even a distributed data processing environment). The terms "comprise," "include," or any other variations thereof are intended to cover non-exclusive inclusion, so that a process, method, product, or device including a series of elements not only includes the elements, but also includes other elements not clearly listed, or further includes elements inherent to the process, method, product, or device. Without more restrictions, it is not excluded that there are other identical or equivalent elements in the process, method, product, or device including the elements. Words such as first and second are used to indicate names, and do not indicate any particular order.

[219] Those skilled in the art also know that, in addition to implementing a controller in the form of pure computer-readable program code, the method steps may absolutely be logically programmed to enable the controller to realize the same function in the form of a logic gate, a switch, a special integrated circuit, a programmable logic controller, an embedded micro-controller, or the like. Therefore, such a controller can be regarded as a hardware component, and apparatuses included therein for realizing various functions can also be regarded as a structure within the hardware component. Or, apparatuses for realizing various functions can even be

regarded as both a software module for implementing a method and a structure within a hardware component.

[220] The present specification may be described in a general context of computer-executable instructions executed by a computer, such as a program module. Generally, the program module includes a routine, a program, an object, an assembly, a data structure, a class, and the like used for executing a specific task or implementing a specific abstract data type. The present specification may also be practiced in distributed computing environments. In the distributed computing environments, tasks are executed by remote processing devices that are linked through a communication network. In the distributed computing environments, the program module may be located in local and remote computer storage media including a storage device.

[221] As can be seen from the description of the above implementations, those skilled in the art can clearly understand that the present specification may be implemented by software plus a necessary universal hardware platform. Based on such an understanding, the technical solutions of the present specification essentially or the part making contributions to the prior art can be embodied in the form of a software product. The computer software product may be stored in a storage medium (such as a ROM/RAM, a magnetic disk, or an optical disc), and include several instructions used for causing a computer device (which may be a personal computer, a mobile terminal, a server, a network device, or the like) to execute the methods in various embodiments or some parts of the embodiments of the present specification.

[222] The embodiments in the present specification are described in a progressive manner, and for the same or similar parts between the embodiments, reference can be made to each other. Each embodiment focuses on the differences from other embodiments. The present specification may be used in many general-purpose or special-purpose computer system environments or configurations, for example: a personal computer, a server computer, a handheld device or portable device, a tablet device, a multiprocessor system, a microprocessor-based system, a set-top box, a

programmable electronic device, a network PC, a minicomputer, a mainframe computer, a distributed computing environment including any of the above systems or devices.

[223] Although the present specification has been described with reference to the embodiments, those of ordinary skills in the art know that the present specification has many variations and changes without departing from the spirit of the present specification, and it is intended to cover in the appended claims these variations and changes without departing from the spirit of the present specification.

CLAIMS

1. A method for constructing a data processing model, comprising:

acquiring model description parameters and sample data of a target data processing model, wherein the model description parameters are provided by a user for the target data processing model and comprise a name of the target data processing model, a data label obtainable by the target data processing model, a data feature useable by the target data processing model, and an evaluation metric of the target data processing model;

determining a base model according to the model description parameters and the sample data; and

training the base model according to the sample data to obtain the target data processing model,

wherein determining the base model according to the model description parameters and the sample data comprises:

determining a model type of the base model from a preset model type group according to the data label, wherein the preset model type group comprises at least a classification type and a regression prediction type;

selecting a plurality of models from models meeting the model type of the base model to serve as a plurality of candidate models;

establishing a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature, and determining a model parameter of each pre-trained model;

determining an adaptability of each pre-trained model according to the evaluation metric, wherein the adaptability of a pre-trained model is configured to represent a degree of applicability of data processing on the sample data based on a pre-trained model and a model parameter corresponding to the pre-trained model; and

determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models.

2. The method of claim 1, wherein selecting a plurality of models from models meeting the model type of the base model to serve as candidate models comprises:

processing the sample data by using data processing models of which applied base models meet the model type of the base model in a preset data processing model library to obtain data processing results of the data processing models;

conducting statistics on evaluation metrics of the data processing models according to the data processing results of the data processing models;

determining a data processing model in which a degree of difference between its evaluation metric and the evaluation metric of the target data processing model is less than a preset degree threshold as a recommended model; and

determining a base model to which the recommended model is applied as a candidate model.

3. The method of claim 1, wherein after determining a model type of the base model from a preset model type group according to the data label, the method further comprises:

determining a data type of the sample data, wherein the data type of the sample data comprises at least one of the following: picture data and text data; and

selecting, according to the data type of the sample data and through filtering the model types of base models, a model type matching the data type of the sample data to serve as the model type of the base model.

4. The method of claim 1, wherein establishing a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature comprises:

extracting a preset proportion of sample data from the sample data to serve as test data;

pre-processing the test data to obtain processed test data, wherein the pre-processing comprises: dimension reduction processing and/or cleaning processing;

selecting feature data from the processed test data according to the data feature; and

training each of the plurality of candidate models respectively according to the feature data to obtain the pre-trained model.

5. The method of claim 1, wherein determining an adaptability of each pre-trained model according to the evaluation metric comprises:

performing a parameter search and a structure search according to the pre-trained model and a corresponding model parameter to obtain a search result; and

determining the adaptability of the pre-trained model according to the search result and the evaluation metric.

6. The method of claim 5, wherein performing a parameter search and a structure search according to the pre-trained model and a corresponding model parameter to obtain a search result comprises:

performing the parameter search and the structure search through at least one of the following algorithms: a Monte Carlo algorithm, a grid algorithm, a Bayesian algorithm, and a genetic algorithm.

7. The method of claim 1, wherein determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models comprises:

determining a model used by the pre-trained model with the highest adaptability as the base model.

8. The method of claim 7, wherein determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models further comprises:

combining models used by the plurality of pre-trained models whose adaptabilities are greater than a preset adaptability threshold to obtain a combined model; and determining the combined model as the base model.

9. The method of claim 1, wherein the model description parameter further comprises: a model type of the base model, and a split rule of the sample data.

10. A data processing method, comprising:

acquiring to-be-processed target data; and

invoking a target data processing model to process the target data, wherein the target data processing model is obtained by determining a base model according to model description parameters and sample data of the target data processing model that are acquired and then training the base model according to the sample data,

wherein the model description parameters are provided by a user for the target data processing model and comprise a name of the target data processing model, a data label obtainable by the target data processing model, a data feature useable by the target data processing model, and an evaluation metric of the target data processing model, and

wherein the base model is determined based on the following:

determining a model type of the base model from a preset model type group according to the data label, wherein the preset model type group comprises at least a classification type and a regression prediction type;

selecting a plurality of models from models meeting the model type of the base model to serve as a plurality of candidate models;

establishing a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature, and determining a model parameter of each pre-trained model;

determining an adaptability of each pre-trained model according to the evaluation metric, wherein the adaptability of a pre-trained model is configured to represent a degree of applicability of data processing on the sample data based on a pre-trained model and a model parameter corresponding to the pre-trained model; and

determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models.

11. A data processing method, comprising:

displaying a data input interface; and

receiving model description parameters and sample data of a target data processing model input by a user on the data input interface, wherein the model description parameters and the sample data are configured to determine a base model, and the base model is configured to establish the target data processing model,

wherein the model description parameters comprise a name of the target data processing model, a data label obtainable by the target data processing model, a data feature useable by the target data processing model, and an evaluation metric of the target data processing model, and

wherein the base model is determined based on the following:

determining a model type of the base model from a preset model type group according to the data label, wherein the preset model type group comprises at least a classification type and a regression prediction type;

selecting a plurality of models from models meeting the model type of the base model to serve as a plurality of candidate models;

establishing a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature, and determining a model parameter of each pre-trained model;

determining an adaptability of each pre-trained model according to the evaluation metric, wherein the adaptability of a pre-trained model is configured to represent a degree of applicability of data processing on the sample data based on a pre-trained model and a model parameter corresponding to the pre-trained model; and

determining the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models.

12. The method of claim 11, wherein after receiving a model description parameter and sample data of a target data processing model input by a user on the data input interface, the method further comprises:

displaying, in response to a display operation, an execution file generated based on the model description parameter; and

receiving a modification operation on the execution file, and updating the execution file according to the modification operation.

13. An apparatus for constructing a data processing model, comprising:

an acquisition module configured to acquire model description parameters and sample data of a target data processing model, wherein the model description parameters are provided by a user for the target data processing model and comprise a name of the target data processing model, a data label obtainable by the target data processing model, a data feature useable by the target data processing model, and an evaluation metric of the target data processing model;

a determination module configured to determine a base model according to the model description parameter and the sample data; and

a training module configured to train the base model according to the sample data to obtain the target data processing model,

wherein the determination module comprises:

a first determination unit configured to determine a model type of the base model from a preset model type group according to the data label, wherein the preset model type group comprises at least a classification type and a regression prediction type;

a selection unit configured to select a plurality of models from models meeting the model type of the base model to serve as candidate models;

an establishment unit configured to establish a pre-trained model for each of the plurality of candidate models according to the sample data and the data feature, and determine a model parameter of each pre-trained model;

a second determination unit configured to determine an adaptability of each pre-trained model according to the evaluation metric, wherein the adaptability is configured to represent a degree of

applicability of data processing on the sample data based on a pre-trained model and a model parameter corresponding to the pre-trained model; and

a third determination unit configured to determine the base model from the plurality of candidate models according to the adaptabilities of the pre-trained models.

14. The apparatus of claim 13, wherein the selection unit comprises:

a processing sub-unit configured to process the sample data by using data processing models of which applied base models meet the model type of the base model in a preset data processing model library to obtain data processing results of the data processing models;

a statistics sub-unit configured to conduct statistics on evaluation metrics of the data processing models according to the data processing results of the data processing models;

a first determination sub-unit configured to determine a data processing model in which a degree of difference between its evaluation metric and the evaluation metric of the target data processing model is less than a preset degree threshold as a recommended model; and

a second determination sub-unit configured to determine a base model to which the recommended model is applied as a candidate model.

15. The apparatus of claim 13, wherein the establishment unit comprises:

an extraction sub-unit configured to extract a preset proportion of sample data from the sample data to serve as test data;

a pre-processing sub-unit configured to pre-process the test data to obtain processed test data, wherein the pre-processing comprises: dimension reduction processing and/or cleaning processing;

a selection sub-unit configured to select feature data from the processed test data according to the data feature; and

a training sub-unit configured to train each of the plurality of candidate models respectively according to the feature data to obtain the pre-trained model.

16. The apparatus of claim 13, wherein the third determination unit comprises: a third determination sub-unit configured to determine a model used by the pre-trained model with the highest adaptability as the base model.

17. The apparatus of claim 16, wherein the third determination unit further comprises: a fourth determination sub-unit configured to combine models used by the plurality of pre-trained models whose adaptabilities are greater than a preset adaptability threshold to obtain a combined model; and determine the combined model as the base model.

18. The apparatus of claim 13, wherein the model description parameter further comprises: a model type of the base model and a split rule of the sample data.

19. A server, comprising a processor and a memory for storing processor-executable instructions, wherein the steps of the method of any one of claims 1 to 9 are implemented when the processor executes the instructions.

20. A user terminal, comprising a processor and a memory for storing processor-executable instructions, wherein the steps of the method of any one of claims 10 to 12 are implemented when the processor executes the instructions.

21. A computer-readable storage medium, storing computer instructions thereon, wherein the steps of the method of any one of claims 1 to 9 are implemented when the instructions are executed.