

WHAT IS CLAIMED IS:

1. A system comprising at least one computing device and at least one storage device storing instructions, wherein the instructions, when executed by the at least one computing device, cause the at least one computing device to perform following steps for automatically training a machine learning model, the steps comprising:

acquiring a data set comprising multiple data records by a data set acquisition device, wherein the data set relates to at least one of: image data, voice data, text data and service data;

determining a feature engineering strategy for the data set through meta-learning by a feature engineering strategy determination device;

generating features related to each data record in the data set using the determined feature engineering strategy by a feature generation device;

constructing a machine learning sample at least based on the generated features by a sample construction device;

training the machine learning model using the constructed machine learning sample by a training device, based on automatically obtained hyper-parameter values; and

performing predicting operation using the the machine learning model, wherein the predicting operation comprises at least one of: predicting image category, predicting text category, predicting speech emotion, predicting fraudulent transaction, and predicting advertising click-through rate,

wherein the determining of the feature engineering strategy for the data set through the meta-learning comprises:

analyzing the data set to obtain a meta feature that reflects overall characteristics of the data set; and

determining the feature engineering strategy for the data set based on the obtained meta feature;

wherein the training of the machine learning model using the constructed machine learning sample based on the automatically obtained hyper-parameter values comprises:

using a part of all the constructed machine learning samples as a training sample set;

generating multiple sets of hyper-parameter values of hyper-parameters that affect the training of the machine learning model by automatically using different hyper-parameter generation strategies according to a number of training samples comprised in the training sample set;

training multiple machine learning models by using the constructed machine learning sample based on each set of hyper-parameter values among the multiple sets of hyper-parameter values, respectively; and

selecting several machine learning models with top-ranked model effects among the multiple trained machine learning models, and fusing the several selected machine learning models through a predetermined model fusion strategy to determine a finally trained machine learning model,

wherein the determining of the finally trained machine learning model comprises:

using the several machine learning models as an initial model candidate set; and

for the initial model candidate set, performing following model combination operations: combining a machine learning model with a best model effect among the several machine learning models with machine learning models in the initial model candidate set, respectively, and calculating the model effect of a respective combined machine learning model group; determining whether the model effect of the respective combined machine learning model group is improved compared with the best model effect before the combination; in response to the model effect of the respective combined machine learning model group being not improved compared with the best model effect before the combination, determining the

machine learning model with the best model effect before the combination as a final machine learning model; in response to the model effect of the respective combined machine learning model group being improved compared with the best model effect before the combination, adding the combined machine learning model group with the best model effect after the combination to the initial model candidate set to constitute a current model candidate set;

repeating, for the current model candidate set, model combination operations the same as the model combination operations performed for the initial model candidate set for a predetermined number of times, and selecting a machine learning model combination with the best model effect in the finally generated current model candidate set; and

determining respective weights of the several machine learning models when the models are fused, according to the number of each machine learning model in the several machine learning models comprised in the machine learning model combination.

2. The system of claim 1, wherein the determining of the feature engineering strategy for the data set based on the obtained meta feature comprises:

automatically determining the feature engineering strategy for the data set based on the obtained meta feature according to correspondence relationship between the meta feature and the feature engineering strategy; or

determining the feature engineering strategy for the data set using another machine learning model based on the obtained meta feature.

3. The system of claim 2, wherein the determining of the feature engineering strategy for the data set using another machine learning model comprises:

determining a historical data set that is similar to the data set and has a corresponding model training result using a clustering model as the another machine learning model based on the obtained meta feature, and determining a feature engineering strategy corresponding to the historical data set as the feature engineering strategy for the data set; or

inputting a prediction sample constituted based on the obtained meta feature to a prediction model as the another machine learning model, and using a prediction result of the prediction model for the prediction sample as the feature engineering strategy for the data set, wherein the prediction model is pre-trained to predict the feature engineering strategy corresponding to the meta feature based on the meta feature.

4. The system of claim 1, wherein the analyzing of the data set to obtain the meta feature that reflects the overall characteristics of the data set comprises:

analyzing the data set to obtain statistic information regarding at least one of a number of data records comprised in the data set, types of features comprised in the data record, a number of the features, and values of the features, as the meta feature.

5. The system of claim 4, wherein the meta feature comprises at least one of: the number of the data records comprised in the data set, a number of discrete features comprised in the data record, a number of consecutive features comprised in the data record, whether there is a time feature in the data record, a number of time features comprised in the data record, a filling degree of each discrete feature, a filling degree of each consecutive feature, an average filling degree of all the discrete features, an average filling degree of all the consecutive features, a mean, variance, median, maximum or minimum value of respective consecutive feature values, and a number of different values of respective discrete features.

6. The system of claim 5, wherein the determining of the feature engineering strategy for the data set based on the obtained meta feature comprises at least one of following operations:

in response to the discrete features being stronger than the consecutive features in distribution, determining the feature engineering strategy for the data set as: discretizing the consecutive features to obtain discretization features, and searching Cartesian product

combinations of original discrete features and the discretization features for a discrete feature combination that is beneficial to a training effect of the model, as the generated feature related to each data record;

in response to the consecutive features being stronger than the discrete features in distribution, determining the feature engineering strategy for the data set as: training a gradient boosting decision tree model using the data set, and using a serial number of a leaf node of each decision tree where the sample constituted based on each data record falls as the generated feature related to each data record;

in response to there being the time feature in the data set, determining the feature engineering strategy for the data set as: sorting the multiple data records according to the time feature, and in a process of training a logistic regression model sequentially using the sorted data records, acquiring a sequence feature corresponding to the current data record based on a parameter item related to a model parameter of the logistic regression model as the generated feature related to each data record; and

in response to the data set comprising multiple time features, determining the feature engineering strategy for the data set as: generating, for each data record, the feature related to each data record by combining the multiple time features in pairs and then subtracting them.

7. The system of claim 1, wherein the generating of multiple sets of hyper-parameter values of the hyper-parameters that affect the training of the machine learning model by automatically using the different hyper-parameter generation strategies according to the number of training samples comprised in the training sample set comprises:

in response to the number of the training samples comprised in the training sample set being less than or equal to a preset threshold, generating the multiple sets of hyper-parameter values of the hyper-parameters by: randomly generating the multiple sets of hyper-parameter values of the hyper-parameters within a range of the hyper-parameter values corresponding to the number of the training samples, or pre-defining hyper-parameter values of partial hyper-parameters among the hyper-parameters, and randomly generating, for remaining hyper-parameters, multiple sets of hyper-parameter values of the remaining hyper-parameters within a range of the hyper-parameter values corresponding to the number of the training samples; and

in response to the number of the training samples comprised in the training sample set being greater than the preset threshold, generating the multiple sets of hyper-parameter values of the hyper-parameters by: pre-defining hyper-parameter values of partial hyper-parameters among the hyper-parameters, and generating, for remaining hyper-parameters, multiple sets of hyper-parameter values of the remaining hyper-parameters by a combination method in which a method of randomly generating the hyper-parameter values within the range of the hyper-parameter values corresponding to the number of the training samples and a method of generating hyper-parameter values using a Bayesian optimization model are combined, wherein the Bayesian optimization model is constructed to reflect relationship between a hyper-parameter value and a model effect corresponding thereto.

8. The system of claim 7, wherein the generating, for the remaining hyper-parameters, of the multiple sets of hyper-parameter values of the remaining hyper-parameters by the combination method comprises:

constructing a sub training sample set by sampling the training samples in the training sample set;

randomly generating, for the remaining hyper-parameters, predetermined sets of initial hyper-parameter values of the remaining hyper-parameters within a range of the hyper-parameter values corresponding to a number of the training samples comprised in the sub training sample set, train the machine learning model using the sub training sample set

based on each set of hyper-parameter values constituted by the pre-defined values of the partial hyper-parameters and each set of the initial hyper-parameter values of the remaining hyper-parameters, and verifying a model effect of the trained machine learning model to obtain a model effect corresponding to the each set of hyper-parameter values;

establishing the Bayesian optimization model according to the each set of hyper-parameter values and the model effect corresponding thereto; and

continuously determining a number of sets of hyper-parameter values that needs to be generated currently for the remaining hyper-parameters, generating one part of the number of sets using the Bayesian optimization model, and randomly generating another part of the number of sets within the range of the hyper-parameter values corresponding to the number of the training samples comprised in the sub training sample set.

9. The system of claim 8, wherein the constructing of the sub training sample set by sampling the training samples in the training sample set comprises: performing multiple times of sampling for the training samples in the training sample set to construct multiple sub training sample sets with different numbers of training samples, wherein among the multiple times of sampling, the sub training sample set obtained from latter sampling comprises more training samples than the sub training sample set obtained from previous sampling; and

wherein the generating, for the remaining hyper-parameters, of the multiple sets of hyper-parameter values of the remaining hyper-parameters by the combination method further comprises:

whenever new hyper-parameter values of the remaining hyper-parameters are generated according to the combination method, training the machine learning model using the sub training sample set obtained from first sampling, based on each set of hyper-parameter values constituted by the pre-defined hyper-parameter values of the partial hyper-parameters and the generated new hyper-parameter values;

selecting multiple sets of hyper-parameter values corresponding to top-ranked model effects among all the sets of hyper-parameter values used when training the machine learning model using the sub training sample set obtained from the previous sampling, as all sets of hyper-parameter values to be used when training the machine learning model using the sub training sample set obtained from next sampling; and

in response to a predetermined condition being satisfied, stopping sampling, and using multiple sets of hyper-parameter values corresponding to top-ranked model effects among all the sets of hyper-parameter values used when training the machine learning model using the sub training sample set obtained from the last sampling, as all sets of hyper-parameter values of the hyper-parameters that affect the training of the machine learning model.

10. The system of claim 8, wherein the constructing of the sub training sample set comprises: constructing the sub training sample set by performing sub-sampling based on density for the training samples in the training sample set.

11. The system of claim 10, wherein the constructing of the sub training sample set by performing the sub-sampling based on density for the training samples in the training sample set comprises:

reducing dimensionality of the training samples in the training sample set to another dimension space; and

constructing the sub training sample set by sampling the training samples according to labels of the training samples and distribution density of the training samples in the training sample set in the another dimension space.

12. The system of claim 11, wherein the reducing of the dimensionality of the training samples in the training sample set to the another dimension space comprises: mapping each training sample in the training sample set to a sample point on a two-dimensional plane; and

wherein the constructing of the sub training sample set by sampling the training samples

according to the labels of the training samples and the distribution density of the training samples in the training sample set in the another dimension space comprises:

- classifying all sample points on the two-dimensional plane into positive sample point sets and negative sample point sets according to whether the training sample is a positive sample or a negative sample; and

- regarding each point set among the positive sample point sets and the negative sample point sets, sampling the sample points in accordance with a set sampling ratio according to distribution density of the sample points in the point set, and using a set constituted by training samples corresponding to sampled sample points as the sub training sample set.

13. The system of claim 12, wherein the sampling of the sample points in accordance with the set sampling ratio according to the distribution density of the sample points in the point set comprises:

- determining a planar region covering all the sample points in the point set, according to a maximum value and a minimum value of abscissas and ordinates of all the sample points in the point set on the two-dimensional plane;

- dividing the determined planar region into multiple sub-regions;

- counting a number of sub-regions where the sample points fall;

- in response to the number of sub-regions where the sample points fall being greater than a preset number, randomly sampling in accordance with the set sampling ratio in each sub-region; and

- in response to the number of sub-regions where the sample points fall being less than or equal to the preset number, increasing the number of the divided sub-regions until the number of sub-regions where the sample points fall is greater than the preset number, and then randomly sampling in accordance with the set sampling ratio in each sub-region.

14. The system of claim 8, wherein the generating, for the remaining hyper-parameters, of the multiple sets of hyper-parameter values of the remaining hyper-parameters by the combination method further comprises:

- after the Bayesian optimization model is established, whenever a new hyper-parameter value and the model effect corresponding thereto are obtained, using them to update the Bayesian optimization model.

15. The system of claim 8, wherein in the case where the number of the training samples in the training sample set is greater than the preset threshold, the determining of the finally trained machine learning model comprises:

- acquiring several sets of hyper-parameter values used when training the several machine learning models with the top-ranked model effects;

- performing secondary training on the machine learning models using the training sample set based on each set of hyper-parameter values of the several sets of hyper-parameter values; and

- fusing the several machine learning models on which the secondary training is performed through the predetermined model fusion strategy to determine the finally trained machine learning model.

16. The system of claim 1, wherein the determining of the finally trained machine learning model further comprises:

- acquiring several sets of hyper-parameter values used when training the several machine learning models with the top-ranked model effects, and a number of rounds of training the machine learning model using the training sample set based on each set of hyper-parameter values of the several sets of hyper-parameter values, respectively;

- training the machine learning model for the corresponding number of rounds using all the constructed machine learning samples based on each set of hyper-parameter values, to

obtain several re-trained machine learning models; and

fusing the several re-trained machine learning models according to the determined respective weights when the models are fused, to obtain the finally trained machine learning model.

17. The system of claim 6, wherein the machine learning model is a neural network model or a logistic regression model.

18. The system of claim 8, wherein the generating of the one part of the number of sets using the Bayesian optimization model and the randomly generating of the another part of the number of sets are performed in parallel.

19. The system of any one of claims 1 to 18, wherein the data set relates to at least one of image data, voice data, text data and service data.

20. A method, executed by at least one computing device, for automatically training a machine learning model comprises:

acquiring a data set comprising multiple data records by a data set acquisition device, wherein the data set relates to at least one of: image data, voice data, text data and service data;

determining a feature engineering strategy for the data set through meta-learning by a feature engineering strategy determination device;

generating features related to each data record in the data set using the determined feature engineering strategy by a feature generation device;

constructing a machine learning sample at least based on the generated features by a sample construction device;

train the machine learning model using the constructed machine learning sample based on automatically obtained hyper-parameter values by a training device; and

performing predicting operation using the the machine learning model, wherein the predicting operation comprises at least one of: predicting image category, predicting text category, predicting speech emotion, predicting fraudulent transaction, and predicting advertising click-through rate,

wherein the determining of the feature engineering strategy for the data set through the meta-learning comprises:

analyzing the data set to obtain a meta feature that reflects overall characteristics of the data set; and

determining the feature engineering strategy for the data set based on the obtained meta feature;

wherein the training of the machine learning model using the constructed machine learning sample based on the automatically obtained hyper-parameter values comprises:

using a part of all the constructed machine learning samples as a training sample set;

generating multiple sets of hyper-parameter values of hyper-parameters that affect the training of the machine learning model by automatically using different hyper-parameter generation strategies according to a number of training samples comprised in the training sample set;

training multiple machine learning models by using the constructed machine learning sample based on each set of hyper-parameter values among the multiple sets of hyper-parameter values, respectively; and

selecting several machine learning models with top-ranked model effects among the multiple trained machine learning models, and fusing the several selected machine learning models through a predetermined model fusion strategy to determine a finally trained machine learning model,

wherein the determining of the finally trained machine learning model comprises:

using the several machine learning models as an initial model candidate set; and

for the initial model candidate set, performing following model combination operations: combining a machine learning model with a best model effect among the several machine learning models with machine learning models in the initial model candidate set, respectively, and calculating the model effect of a respective combined machine learning model group; determining whether the model effect of the respective combined machine learning model group is improved compared with the best model effect before the combination; in response to the model effect of the respective combined machine learning model group being not improved compared with the best model effect before the combination, determining the machine learning model with the best model effect before the combination as a final machine learning model; in response to the model effect of the respective combined machine learning model group being improved compared with the best model effect before the combination, adding the combined machine learning model group with the best model effect after the combination to the initial model candidate set to constitute a current model candidate set;

repeating, for the current model candidate set, model combination operations the same as the model combination operations performed for the initial model candidate set for a predetermined number of times, and selecting a machine learning model combination with the best model effect in the finally generated current model candidate set; and

determining respective weights of the several machine learning models when the models are fused, according to the number of each machine learning model in the several machine learning models comprised in the machine learning model combination.

21. The method of claim 20, wherein the determining of the feature engineering strategy for the data set based on the obtained meta feature comprises:

automatically determining the feature engineering strategy for the data set based on the obtained meta feature according to correspondence relationship between the meta feature and the feature engineering strategy; or

determining the feature engineering strategy for the data set using another machine learning model based on the obtained meta feature.

22. The method of claim 21, wherein the determining of the feature engineering strategy for the data set using the another machine learning model comprises:

determining a historical data set that is similar to the data set and has a corresponding model training result using a clustering model as the another machine learning model based on the obtained meta feature, and determining a feature engineering strategy corresponding to the historical data set as the feature engineering strategy for the data set; or

inputting a prediction sample constituted based on the obtained meta feature to a prediction model as the another machine learning model, and using a prediction result of the prediction model for the prediction sample as the feature engineering strategy for the data set, wherein the prediction model is pre-trained to predict the feature engineering strategy corresponding to the meta feature based on the meta feature.

23. The method of claim 20, wherein the analyzing of the data set to obtain the meta feature that reflects the overall characteristics of the data set comprises:

analyzing the data set to obtain statistic information regarding at least one of a number of data records comprised in the data set, types of features comprised in the data record, a number of the features, and values of the features, as the meta feature.

24. The method of claim 23, wherein the meta feature comprises at least one of: the number of the data records comprised in the data set, a number of discrete features comprised in the data record, a number of consecutive features comprised in the data record, whether there is a time feature in the data record, a number of time features comprised in the data record, a filling degree of each discrete feature, a filling degree of each consecutive feature, an average filling degree of all the discrete features, an average filling degree of all the consecutive features, a mean, variance, median, maximum or minimum value of respective

consecutive feature values, and a number of different values of respective discrete features.

25. The method of claim 24, wherein the determining of the feature engineering strategy for the data set based on the obtained meta feature comprises at least one of following operations:

in response to the discrete features being stronger than the consecutive features in distribution, determining the feature engineering strategy for the data set as: discretizing the consecutive features to obtain discretization features, and searching Cartesian product combinations of original discrete features and the discretization features for a discrete feature combination that is beneficial to a training effect of the model, as the generated feature related to each data record;

in response to the consecutive features being stronger than the discrete features in distribution, determining the feature engineering strategy for the data set as: training a gradient boosting decision tree model using the data set, and using a serial number of a leaf node of each decision tree where the sample constituted based on each data record falls as the generated feature related to each data record;

in response to there being the time feature in the data set, determining the feature engineering strategy for the data set as: sorting the multiple data records according to the time feature, and in a process of training a logistic regression model sequentially using the sorted data records, acquiring a sequence feature corresponding to the current data record based on a parameter item related to a model parameter of the logistic regression model as the generated feature related to each data record; and

in response to the data set comprising multiple time features, determining the feature engineering strategy for the data set as: generating, for each data record, the feature related to each data record by combing the multiple time features in pairs and then subtracting them.

26. The method of claim 19, wherein the generating of multiple sets of hyper-parameter values of the hyper-parameters that affect the training of the machine learning model by automatically using the different hyper-parameter generation strategies according to the number of training samples comprised in the training sample set comprises:

in response to the number of the training samples comprised in the training sample set being less than or equal to a preset threshold, generating the multiple sets of hyper-parameter values of the hyper-parameters by: randomly generating the multiple sets of hyper-parameter values of the hyper-parameters within a range of the hyper-parameter values corresponding to the number of the training samples, or pre-defining hyper-parameter values of partial hyper-parameters among the hyper-parameters, and randomly generating multiple sets of hyper-parameter values of remaining hyper-parameters within the range of the hyper-parameter values corresponding to the number of the training samples for the remaining hyper-parameters; and

in response to the number of the training samples comprised in the training sample set being greater than the preset threshold, generating the multiple sets of hyper-parameter values of the hyper-parameters by: pre-defining hyper-parameter values of partial hyper-parameters among the hyper-parameters, and generating, for remaining hyper-parameters, multiple sets of hyper-parameter values of the remaining hyper-parameters by a combination method in which a method of randomly generating the hyper-parameter values within the range of the hyper-parameter values corresponding to the number of the training samples and a method of generating hyper-parameter values using a Bayesian optimization model are combined, wherein the Bayesian optimization model is constructed to reflect relationship between a hyper-parameter value and a model effect corresponding thereto.

27. The method of claim 26, wherein the generating, for the remaining hyper-parameters, of the multiple sets of hyper-parameter values of the remaining hyper-parameters by the

combination method comprises:

- constructing a sub training sample set by sampling the training samples in the training sample set;

- randomly generating, for the remaining hyper-parameters, predetermined sets of initial hyper-parameter values of the remaining hyper-parameters within the range of the hyper-parameter values corresponding to a number of training samples comprised in the sub training sample set, train the machine learning model using the sub training sample set based on each set of hyper-parameter values constituted by the pre-defined values of the partial hyper-parameters and each set of the initial hyper-parameter values of the remaining hyper-parameters, and verifying a model effect of the trained machine learning model to obtain a model effect corresponding to the each set of hyper-parameter values;

- establishing the Bayesian optimization model according to the each set of hyper-parameter values and the model effect corresponding thereto; and

- continuously determining a number of sets of hyper-parameter values that needs to be generated currently for the remaining hyper-parameters, generating one part of the number of sets using the Bayesian optimization model, and randomly generating another part of the number of sets within the range of the hyper-parameter values corresponding to the number of the training samples comprised in the sub training sample set.

28. The method of claim 27, wherein the constructing of the sub training sample set by sampling the training samples in the training sample set comprises: performing multiple times of sampling for the training samples in the training sample set to construct multiple sub training sample sets with different numbers of training samples, wherein among the multiple times of sampling, the sub training sample set obtained from the latter sampling comprises more training samples than the sub training sample set obtained from the previous sampling; and

wherein the generating, for the remaining hyper-parameters, of the multiple sets of hyper-parameter values of the remaining hyper-parameters by the combination method further comprises:

- whenever new hyper-parameter values of the remaining hyper-parameters are generated according to the combination method, training the machine learning model using the sub training sample set obtained from the first sampling, based on each set of hyper-parameter values constituted by the pre-defined hyper-parameter values of the partial hyper-parameters and the generated new hyper-parameter values;

- selecting multiple sets of hyper-parameter values corresponding to top-ranked model effects among all the sets of hyper-parameter values used when training the machine learning model using the sub training sample set obtained from the previous sampling, as all sets of hyper-parameter values to be used when training the machine learning model using the sub training sample set obtained from the next sampling; and

- in response to a predetermined condition being satisfied, stopping sampling, and using multiple sets of hyper-parameter values corresponding to top-ranked model effects among all the sets of hyper-parameter values used when training the machine learning model using the sub training sample set obtained from the last sampling, as all sets of hyper-parameter values of the hyper-parameters that affect the training of the machine learning model.

29. The method of claim 27, wherein the constructing of the sub training sample set comprises: constructing the sub training sample set by performing sub-sampling based on density for the training samples in the training sample set.

30. The method of claim 29, wherein the constructing of the sub training sample set by performing sub-sampling based on density for the training samples in the training sample set comprises:

- reducing dimensionality of the training samples in the training sample set to another

dimension space; and

constructing the sub training sample set by sampling the training samples according to labels of the training samples and distribution density of the training samples in the training sample set in the another dimension space.

31. The method of claim 30, wherein the reducing of the dimensionality of the training samples in the training sample set to the another dimension space comprises: mapping each training sample in the training sample set to a sample point on a two-dimensional plane; and

wherein the constructing of the sub training sample set by sampling the training samples according to the labels of the training samples and the distribution density of the training samples in the training sample set in the another dimension space comprises:

classifying all sample points on the two-dimensional plane into positive sample point sets and negative sample point sets according to whether the training sample is a positive sample or a negative sample; and

regarding each point set among the positive sample point sets and the negative sample point sets, sampling the sample points in accordance with a set sampling ratio according to distribution density of the sample points in the point set, and using a set constituted by training samples corresponding to the sampled sample points as the sub training sample set.

32. The method of claim 31, wherein the sampling of the sample points in accordance with the set sampling ratio according to the distribution density of the sample points in the point set comprises:

determining a planar region covering all the sample points in the point set, according to a maximum value and a minimum value of abscissas and ordinates of all the sample points in the point set on the two-dimensional plane;

dividing the determined planar region into multiple sub-regions;

counting a number of sub-regions where the sample points fall;

in response to the number of sub-regions where the sample points fall being greater than a preset number, randomly sampling in accordance with the set sampling ratio in each sub-region; and

in response to the number of sub-regions where the sample points fall being less than or equal to the preset number, increasing the number of the divided sub-regions until the number of sub-regions where the sample points fall is greater than the preset number, and then randomly sampling in accordance with the set sampling ratio in each sub-region.

33. The method of claim 27, wherein the generating, for the remaining hyper-parameters, of the multiple sets of hyper-parameter values of the remaining hyper-parameters by the combination method further comprises:

after the Bayesian optimization model is established, whenever a new hyper-parameter value and a model effect corresponding thereto are obtained, using them to update the Bayesian optimization model.

34. The method of claim 27, wherein in the case where the number of the training samples in the training sample set is greater than the preset threshold, the determining of the finally trained machine learning model comprises:

acquiring several sets of hyper-parameter values used when training the several machine learning models with the top-ranked model effects;

performing secondary training on the machine learning models using the training sample set based on each set of hyper-parameter values of the several sets of hyper-parameter values; and

fusing the several machine learning models on which the secondary training is performed through the predetermined model fusion strategy to determine the finally trained machine learning model.

35. The method of claim 19, wherein the determining of the finally trained machine learning model further comprises:

- acquiring several sets of hyper-parameter values used when training the several machine learning models with the top-ranked model effects, and a number of rounds of training the machine learning model using the training sample set based on each set of hyper-parameter values of the several sets of hyper-parameter values, respectively;

- training the machine learning model for the corresponding number of rounds using all the constructed machine learning samples based on each set of hyper-parameter values, to obtain several re-trained machine learning models; and

- fusing the several re-trained machine learning models according to the determined respective weights when the models are fused, to obtain the finally trained machine learning model.

36. The method of claim 25, wherein the machine learning model is a neural network model or a logistic regression model.

37. The method of claim 27, wherein the generating of the one part of the number of sets using the Bayesian optimization model and the randomly generating of the another part of the number of sets are performed in parallel.

38. The method of any one of claims 19 to 37, wherein the data set relates to at least one of image data, voice data, text data and service data.

39. A computer readable storage medium storing instructions, wherein the instructions, when executed by at least one computing device, cause the at least one computing device to execute the method of any one of claims 20 to 38.

40. A system for automatically training a machine learning model, comprising:

- a data set acquisition device configured to acquire a data set comprising multiple data records;

- a feature engineering strategy determination device configured to determine a feature engineering strategy for the data set through meta-learning, wherein the data set relates to at least one of: image data, voice data, text data and service data;

- a feature generation device configured to generate features related to each data record in the data set using the determined feature engineering strategy;

- a sample construction device configured to construct a machine learning sample at least based on the generated features;

- a training device configured to train the machine learning model using the constructed machine learning sample, based on automatically obtained hyper-parameter values; and

- performing predicting operation using the the machine learning model, wherein the predicting operation comprises at least one of: predicting image category, predicting text category, predicting speech emotion, predicting fraudulent transaction, and predicting advertising click-through rate,

- wherein the feature engineering strategy determination device is configured to:

- analyze the data set to obtain a meta feature that reflects overall characteristics of the data set; and

- determine the feature engineering strategy for the data set based on the obtained meta feature;

- wherein the training device is configured to:

- use a part of all the constructed machine learning samples as a training sample set;

- generate multiple sets of hyper-parameter values of hyper-parameters that affect the training of the machine learning model by automatically using different hyper-parameter generation strategies according to a number of training samples comprised in the training sample set;

- train multiple machine learning models by using the constructed machine learning

sample based on each set of hyper-parameter values among the multiple sets of hyper-parameter values, respectively; and

select several machine learning models with top-ranked model effects among the multiple trained machine learning models, and fuse the several selected machine learning models through a predetermined model fusion strategy to determine a finally trained machine learning model;

wherein the training device is configured to determine the finally trained machine learning model by:

using the several machine learning models as an initial model candidate set; and

for the initial model candidate set, performing following model combination operations: combining a machine learning model with a best model effect among the several machine learning models with machine learning models in the initial model candidate set, respectively, and calculating the model effect of a respective combined machine learning model group; determining whether the model effect of the respective combined machine learning model group is improved compared with the best model effect before the combination; in response to the model effect of the respective combined machine learning model group being not improved compared with the best model effect before the combination, determining the machine learning model with the best model effect before the combination as a final machine learning model; in response to the model effect of the respective combined machine learning model group being improved compared with the best model effect before the combination, adding the combined machine learning model group with the best model effect after the combination to the initial model candidate set to constitute a current model candidate set;

repeating, for the current model candidate set, model combination operations the same as the model combination operations performed for the initial model candidate set for a predetermined number of times, and selecting a machine learning model combination with the best model effect in the finally generated current model candidate set; and

determining respective weights of the several machine learning models when the models are fused, according to the number of each machine learning model in the several machine learning models comprised in the machine learning model combination.