

DATA PROCESSING METHOD AND DEVICE, MODEL OPTIMIZATION METHOD AND DEVICE

[0001] The present application claims priority to Chinese Patent Application No. 202411212294.4, titled "DATA PROCESSING METHOD AND DEVICE, MODEL OPTIMIZATION METHOD AND DEVICE", filed on August 30, 2024 with the National Intellectual Property Administration, PRC, which is incorporated herein by reference in its entirety.

FIELD

[0002] Embodiments of the present disclosure relate to the technical field of machine learning, and in particular to a data processing method and device, and a model optimization method and device.

BACKGROUND

[0003] With the development of computer and Internet technology, large language models have been significantly developed in the field of natural language processing. A well-trained large language model may support users to conduct personalized Q&A, which can meet usage needs of the users in different scenarios. However, the LLM that relies solely on imitation and learning has a significant problem that such LLM may generate content that appears reasonable but is inaccurate or insecure. In the conventional art, in order to solve this problem, information feedback from humans is usually used to optimize the model, for example, by collecting a large amount of user evaluation data on the model output, and optimizing the model by using a reward model in combination with the evaluation data, to make the generated output of the model meet the requirements of the users. However, this optimization method is overly dependent on feedback content of the users, and the data has small size and insufficient diversity. Therefore, an effective solution is needed urgently to address the above problem.

SUMMARY

[0004] In view of this, a data processing method is provided according to an embodiment of the present disclosure. A data processing device, a model optimization method, a model optimization device, a computing device, a computer-readable storage medium and a computer program product are further provided according to one or more embodiments of the present disclosure to solve the technical deficiencies existing in the conventional art.

[0005] In a first aspect of the embodiments according to the present disclosure, a data processing method is provided, which includes:

performing sampling for multiple datasets, and determining target sample data based on a sampling result;

inputting the target sample data into multiple language models for processing, to obtain target prediction data outputted by each of the language models;

performing text annotation processing on the target prediction data by a preset text annotation model, to obtain text annotation information corresponding to the target prediction data; and

constructing a sample data array based on the target sample data, the target prediction data and the text annotation information, where the sample data array is used to perform a model training task and a model validation task.

[0006] Preferably, the performing sampling for multiple datasets, and determining target sample data based on a sampling result includes:

determining the multiple datasets, where each of the datasets includes sample data of a different data type from other datasets;

performing sampling for each of the datasets to obtain multiple pieces of initial sample data, and constructing an initial sample dataset based on the multiple pieces of initial sample data; and

selecting the target sample data from the initial sample dataset.

[0007] Preferably, before the inputting the target sample data into multiple language models for processing, to obtain target prediction data outputted by each of the language models, the method further includes:

selecting the multiple language models for the target sample data and determining

prompt information corresponding to each of the language models, where

processing the target sample data by any one of the multiple language models includes:

updating the target sample data based on the prompt information corresponding to the language model, and generating model input context information based on an update
5 result; and

inputting the model input context information into the language model for processing, to obtain the target prediction data.

[0008] Preferably, the text annotation information corresponding to the target prediction data outputted by each of the language models is determined by:

10 inputting the target sample data and the target prediction data into a preset text annotation model, processing the target prediction data in at least two annotation dimensions by the text annotation model, and determining, based on a processing result, an annotation text and prediction scores corresponding to the at least two annotation dimensions; and

generating text annotation information corresponding to the target prediction data based
15 on the annotation text and the prediction scores corresponding to the at least two annotation dimensions.

[0009] Preferably, the method further includes:

determining a preset candidate text annotation model, and processing the target prediction data by the candidate text annotation model, to obtain a candidate annotation text
20 corresponding to the target prediction data; and

calculating a text similarity between the candidate annotation text and the annotation text, where

in a case that the text similarity is greater than a preset similarity threshold, the step of generating text annotation information corresponding to the target prediction data based on
25 the annotation text and the prediction scores corresponding to the at least two annotation dimensions is performed.

[0010] Preferably, after the constructing a sample data array based on the target sample data, the target prediction data and the text annotation information, the method further includes:

adding the sample data array to an intermediate sample dataset;

performing data cleaning and format conversion on the intermediate sample dataset in a case that the intermediate sample dataset meets a dataset construction condition; and

generating a target sample dataset based on a data cleaning result and a format conversion result, where the target sample data array in the target sample dataset is used to perform the model training task and the model validation task.

[0011] Preferably, the method further includes:

determining a reward model and a criticism model corresponding to a language model to be optimized;

training the reward model and the criticism model based on the target sample array in the target sample dataset, and determining a target reward model and a target criticism model based on a training result; and

optimizing the language model to be optimized by using the target reward model and the target criticism model until obtaining a target language model that meets an optimization stopping condition.

[0012] In a second aspect of the embodiments according to the present disclosure, a model optimization method is provided, which includes:

determining a reward model and a criticism model corresponding to a language model to be optimized;

training the reward model and the criticism model based on the target sample array in the target sample dataset, and determining a target reward model and a target criticism model based on a training result, where the target sample data array in the target sample dataset is obtained with the above method; and

optimizing the language model to be optimized by using the target reward model and the target criticism model until obtaining a target language model that meets an optimization stopping condition.

[0013] In a third aspect of the embodiments according to the present disclosure, a data processing device is provided, which includes: a determination module, a processing module, an annotation module and a construction module. The determination module is configured to perform sampling for multiple datasets, and determining target sample data based on a

sampling result. The processing module is configured to input the target sample data into multiple language models for processing, to obtain target prediction data outputted by each of the language models. The annotation module is configured to perform text annotation processing on the target prediction data by a preset text annotation model, to obtain text annotation information corresponding to the target prediction data. The construction module is configured to construct a sample data array based on the target sample data, the target prediction data and the text annotation information, where the sample data array is used to perform a model training task and a model validation task.

[0014] In a fourth aspect of the embodiments according to the present disclosure, a model optimization device is provided, which includes: a model determination module, a model training module and a model optimization module. The model determination module is configured to determine a reward model and a criticism model corresponding to a language model to be optimized. The model training module is configured to train the reward model and the criticism model based on the target sample array in the target sample dataset, and determine a target reward model and a target criticism model based on a training result, where the target sample data array in the target sample dataset is obtained with the above method. The model optimization module is configured to optimize the language model to be optimized by using the target reward model and the target criticism model until a target language model that meets an optimization stopping condition is obtained.

[0015] In a fifth aspect of the embodiments according to the present disclosure, a computing device is provided, which includes a memory and a processor. The memory is configured to store computer executable instructions, the processor is configured to execute the computer executable instructions, where the computer executable instructions, when executed by the processor, perform the steps of the data processing method or the model optimization method described above.

[0016] In a sixth aspect of the embodiments according to the present disclosure, a computer-readable storage medium is provided, which includes computer executable instructions stored thereon, where the computer executable instructions, when executed by a processor, perform the steps of the data processing method or the model optimization method described above.

[0017] In a seventh aspect of the embodiments according to the present disclosure, a

computer program product is provided, which includes computer programs or instructions, where the computer programs or instructions, when executed by a processor, perform the steps of the data processing method or the model optimization method described above.

[0018] With the data processing method according to the embodiment, in order to provide abundant and diversified feedback data for model optimization of language models, the feedback data may be constructed by using a text annotation model and multiple language models. Specifically, sampling may be performed for multiple datasets, to determine target sample data based on a sampling result, the sample diversity can be improved by using multiple datasets. Thereafter, the target sample data may be inputted to the multiple language models for processing, and the target prediction data outputted from each of the multiple language models may be obtained. On this basis, text annotation processing may be performed on the target prediction data by a preset text annotation model, to obtain the text annotation information corresponding to the target prediction data. In this way, when constructing the text annotation information, in order to ensure the quality and scale of the data, multiple language models may be used to output multiple pieces of target prediction information, to effectively ensure the abundance of the data. Then, a sample data array may be constructed based on the target sample data, the target prediction data and the text annotation information corresponding to the target prediction data, so that the model training task and the model validation task can be performed in the application stage based on the sample data array, thereby improving the performance and alignment effect of the language model.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] Figure 1 is a schematic diagram of a data processing method according to an embodiment of the present disclosure;

[0020] Figure 2 is a flowchart of a data processing method according to an embodiment of the present disclosure;

[0021] Figure 3 is a flowchart of a model optimization method according to an embodiment of the present disclosure;

[0022] Figure 4 is a flowchart of a processing procedure of a data processing method according to an embodiment of the present disclosure;

[0023] Figure 5 is a schematic diagram of a structure of a data processing device according to an embodiment of the present disclosure;

[0024] Figure 6 is a schematic diagram of a structure of a model optimization device according to an embodiment of the present disclosure; and

5 [0025] Figure 7 is a block diagram of a structure of a computing device according to an embodiment of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

10 [0026] Many specific details are set forth in the following description to facilitate a full understanding of the present disclosure. However, the present disclosure may be implemented in many other ways different from those described herein, and those skilled in the art may make similar generalizations without violating the connotations of the present disclosure, and thus the present disclosure is not limited by the specific embodiments described below.

15 [0027] Terms used in one or more embodiments of the present disclosure are used only for the purpose of describing a particular embodiment and are not intended to limit one or more embodiments of the present disclosure. The singular forms of "a", "the", and "said" used in one or more embodiments of the present disclosure and the claims are also intended to encompass the plural form, unless the context clearly indicates otherwise. It should also be understood that the term "and/or" used in one or more embodiments of the present disclosure
20 refers to and encompasses any or all possible combinations of one or more of the listed items in association.

[0028] It should be understood that while the terms "first", "second", etc. may be used in one or more embodiments of the present disclosure to describe various information, such information should not be limited to these terms. These terms are used only to distinguish the
25 same type of information from one another. For example, without departing from the scope of the one or more embodiments of the present disclosure, "a first..." may also be referred to as "a second...", and similarly, "a second..." may be referred to as "a first...". Depending on the context, the phrase "if" as used herein may be interpreted as "at the time of" or "when" or "in response to a determination".

30 [0029] In addition, it should be noted that the user information (including but not limited to

user device information, user personal information, etc.) and data (including but not limited to data used for analysis, stored data, displayed data, etc.) involved in one or more embodiments of the present disclosure are authorized by the user or sufficiently authorized by the parties, and that the set, use, and processing of the relevant data are subject to relevant laws, regulations, and regulations of relevant countries and regions, and a corresponding operation portal is provided for the user to choose to authorize or reject.

[0030] First, the terminologies involved in one or more embodiments of the present disclosure are explained below.

[0031] The Large Language Model (LLM) is a kind of deep learning model in the field of natural language processing (NLP). This model is known for its large parameter scale, powerful language understanding and generation capabilities. The LLM learn the complex patterns and rules of language by training based on large-scale textual datasets, and thus can perform various NLP tasks, including but not limited to text generation, text categorization, sentiment analysis, question-answer system, and the like.

[0032] In the present disclosure, a data processing method is provided. A data processing device, a model optimization method, a model optimization device, a computing device, a computer-readable storage medium and a computer program product are further provided according to one or more embodiments of the present disclosure, which will be described in detail in the following embodiments.

[0033] In practice, although the language model may be optimized based on feedback data of the user, due to time and budget constraints of the feedback data from the users, it is difficult to scale up user feedback data on a large scale, and there is insufficient diversity. In addition, different users have different subjective consciousness, which may lead to inconsistent annotation standards for feedback data. Therefore, an effective solution is urgently needed to solve the above problem.

[0034] Referring to the schematic diagram shown in Figure 1, with the data processing method according to the embodiment, in order to provide abundant and diversified feedback data for model optimization of language models, the feedback data may be constructed by using a text annotation model and multiple language models. Specifically, sampling may be performed for multiple datasets, to determine target sample data based on a sampling result, the sample diversity can be improved by using multiple datasets. Thereafter, the target sample

data may be inputted to the multiple language models for processing, and the target prediction data outputted from each of the multiple language models may be obtained. On this basis, text annotation processing may be performed on the target prediction data by a preset text annotation model, to obtain the text annotation information corresponding to the target prediction data. In this way, when constructing the text annotation information, in order to ensure the quality and scale of the data, multiple language models may be used to output multiple pieces of target prediction information, to effectively ensure the abundance of the data. Then, a sample data array may be constructed based on the target sample data, the target prediction data and the text annotation information corresponding to the target prediction data, so that the model training task and the model validation task can be performed in the application stage based on the sample data array, thereby improving the performance and alignment effect of the language model.

[0035] Referring to Figure 2, Figure 2 is a flowchart of a data processing method according to an embodiment of the present disclosure, which includes the following steps S202 to S208.

[0036] In step S202, sampling is performed for multiple datasets, and target sample data is determined based on a sampling result.

[0037] The data processing method in this embodiment may be applied to a scenario of constructing feedback data corresponding to a large language model under any business project. The large language model that needs to be optimized for the constructed sample data array may be a LLM model, or other large language models with a natural language processing capability, which is not limited in the embodiment.

[0038] Specifically, multiple datasets refer to sets of sample data provided by different data sources. Each dataset includes the same type of sample data, but the types of data corresponding to the datasets are different, thereby improving the diversity of constructed sample data arrays. Correspondingly, the target sample data specifically refers to any sample data that needs to be processed at the current time after the sampling processing.

[0039] Based on this, in order to provide abundant and diversified feedback data for optimizing language models, feedback data may be constructed by using a text annotation model and multiple language models. Specifically, sampling may be performed for multiple datasets to determine target sample data based on a sampling result, so that the sample diversity can be improved through multiple datasets. Afterwards, the target sample data may

be inputted into multiple language models for processing, to obtain target prediction data outputted by each language model. On this basis, text annotation processing may be performed on the target prediction data by the preset text annotation model, to obtain text annotation information corresponding to the target prediction data. In this way, when constructing the text annotation information, in order to ensure the quality and scale of the data, multiple language models may be used to output multiple pieces of target prediction information, to effectively ensure the abundance of the data. Then, a sample data array may be constructed based on the target sample data, the target prediction data and the text annotation information corresponding to the target prediction data, so that the model training task and the model validation task can be performed in the application stage based on the sample data array.

[0040] Furthermore, in order to construct abundant feedback data to improve the coverage scope of business of the optimized model, a hierarchical sampling way may be used to perform the sampling processing. In this embodiment, the implementation is specifically described as follows:

determining the multiple datasets, where each of the datasets includes sample data of a different data type from other datasets; performing sampling for each of the datasets to obtain multiple pieces of initial sample data, and constructing an initial sample dataset based on the multiple pieces of initial sample data; and selecting the target sample data from the initial sample dataset.

[0041] Based on this, when performing sampling, multiple datasets may be determined first, and the data types of sample data in the datasets are different. Afterwards, sampling may be performed separately for each dataset. During the sampling process, a hierarchical sampling strategy may be used to improve sample abundance. Multiple pieces of initial sample data may be obtained, and an initial sample dataset may be constructed based on the multiple pieces of initial sample data. Subsequently, any sample data may be selected from the initial sample dataset as the target sample data for constructing feedback data.

[0042] In practice, multiple datasets may be selected from TruthfulQA, FalseQA, Evol-Instruct, UltraChat, ShareGPT, etc, which is not limited in this embodiment. When performing sampling, hierarchical sampling may be used to obtain different types of sample data from multiple data sources, ensuring that the sample data can cover multiple tasks and

scenarios. In addition, excessively long or unsuitable sample data may be removed to ensure the simplicity and clarity of the sample data.

[0043] For example, when constructing feedback data for optimizing LLM model, hierarchical sampling may be performed from multiple datasets such as TruthfulQA, FalseQA, Evol-Instruct, UltraChat, ShareGPT, etc. Based on the sampling result, n pieces of sample data may be obtained. Then, by filtering low-quality sample data from the n pieces of sample data, m pieces of target sample data may be obtained, so that feedback data can be constructed separately for the m pieces of target sample data, thereby optimizing the LLM model.

[0044] In practice, the target sample data may be a problem text, an image, a video, etc. inputted to a model for processing, which is not limited in this embodiment.

[0045] In summary, by using the hierarchical sampling way to construct the initial sample dataset, the sample abundance can be effectively improved, thereby covering more task scenarios, thus enhancing the generalization capability of the model.

[0046] In step S204, the target sample data is inputted into multiple language models for processing, to obtain target prediction data outputted by each language model.

[0047] Specifically, after obtaining the target sample data mentioned above, in order to ensure the diversity of feedback data, the target sample data may be processed by multiple language models to obtain the target prediction data outputted by each language model separately, so as to complete construction of the feedback data by multiple pieces of target prediction data.

[0048] The multiple language models specifically refer to language models that may process the target sample data and output prediction results, such as GPT-4, gpt-3.5-turbo, Bard, which is not limited in the embodiment. The number of the multiple language models selected during the construction of the prediction data may be set according to actual needs, and is not limited in the embodiment. Correspondingly, the target prediction data refers to a prediction result obtained by the model after processing the target sample data. For example, if the target sample data is a question text, the target prediction data is an answer text outputted by the language model. Due to the different capabilities of different models, multiple different pieces of prediction data may be obtained.

[0049] Furthermore, in order to output different prediction information through multiple

language models and obtain different text annotation information to improve the abundance of feedback data, the sample data may be updated based on different prompt information corresponding to different language models and prediction information is generated. In this embodiment, the implementation is specifically described as follows:

5 selecting the multiple language models for the target sample data and determining prompt information corresponding to each of the language models, where

 processing the target sample data by any one of the multiple language models includes:

 updating the target sample data based on the prompt information corresponding to the language model, and generating model input context information based on an update result;

10 and

 inputting the model input context information into the language model for processing, to obtain the target prediction data.

[0050] Specifically, the prompt information refers to prompt words corresponding to different language models. Correspondingly, the model input context information refers to
15 text information obtained by updating the target sample data using the prompt information, which is used to input to the model for generating prediction data.

[0051] Based on this, when selecting multiple language models for prediction data generation, multiple language models may be selected based on the target sample data, and the prompt information corresponding to each language model needs to be determined. After
20 updating the target sample data with different prompt information, different language models may provide different prediction data for subsequent construction of diversity feedback data. For the processing of target sample data by any one of the multiple language models, the target sample data may be updated first based on the prompt information corresponding to the language model, and model input context information is generated based on the update result.
25 Afterwards, the model may input the model input context information into the language model for processing, to obtain the target prediction data outputted by the language model. Similarly, any language model may process in this method to obtain multiple pieces of prediction data for subsequent use.

[0052] In practice, in order to avoid the problem of consistency in text style and quality of
30 prediction data generated for sample data, multiple different language models may be

randomly selected to output target prediction data corresponding to the target sample data. Different principles and prompt words may also be introduced to update the sample data to induce the language model to output target prediction data with different styles and qualities.

5 **[0053]** Following the previous example, after obtaining m pieces of target sample data (for example, $m=2$ in this embodiment), two language models may be selected for target sample data 1 and target sample data 2, namely a language model a and a language model b. Then, prompt words corresponding to the language model a and the language model b may be determined, and the target sample data 1 may be updated to target sample data 1a and target sample data 1b, and the target sample data 2 may be updated to target sample data 2a and
10 target sample data 2b. Furthermore, the target sample data 1a and the target sample data 2a are inputted into the language model a for processing, the target prediction data 1a corresponding to the target sample data 1 and the target prediction data 2a corresponding to the target sample data 2 may be obtained. In addition, the target sample data 1b and the target sample data 2b is inputted into the language model b for processing, the target prediction data 1b corresponding
15 to the target sample data 1 and the target prediction data 2b corresponding to the target sample data 2 may be obtained, where each prediction data corresponds to a different style and quality, facilitating the subsequent use.

[0054] In summary, by introducing prompt information to update the sample data, the language model may output different responses to different quality and styles of sample data,
20 thereby realizing subsequent construction of feedback data by using diversified data.

[0055] In step S206, a text annotation processing is performed on each target prediction data by a preset text annotation model, to obtain text annotation information corresponding to each target prediction data.

[0056] Specifically, after obtaining multiple pieces of prediction data corresponding to the
25 target sample data as described above, in order to improve the efficiency of feedback data construction and reduce costs, a text annotation processing is performed on each target prediction data by a preset text annotation model, to obtain text annotation information corresponding to each target prediction data, achieving automated text annotation by the text annotation model, thus improving the efficiency of feedback data construction.

30 **[0057]** The text annotation model specifically refers to a model that may generate a feedback annotation for target prediction data in a case that the target prediction data

corresponding to the target sample data is inputted into this model, which may be implemented by using GPT-4. Correspondingly, the text annotation information specifically refers to annotation information obtained by the text annotation model after performing text annotation processing on the target prediction data, which is used to simulate real feedback data of the user, facilitating subsequent optimization of the language model.

[0058] Furthermore, the text annotation information may be constructed based on the annotation text and the prediction score to facilitate the subsequent selection of high-quality feedback data for use. In this embodiment, the implementation is specifically described as follows:

inputting the target sample data and the target prediction data into a preset text annotation model, processing the target prediction data in at least two annotation dimensions by the text annotation model, and determining, based on a processing result, an annotation text and prediction scores corresponding to the at least two annotation dimensions; and generating text annotation information corresponding to the target prediction data based on the annotation text and the prediction scores corresponding to the at least two annotation dimensions.

[0059] Specifically, the prediction score refers to a score obtained by the text annotation model by scoring for the target prediction data in at least two annotation dimensions. The at least two annotation dimensions may include an instruction compliance dimension, an authenticity dimension, a honesty dimension, and a usefulness dimension, and the availability of the target prediction data is reflected through the corresponding prediction scores of multiple dimensions, so as to facilitate the subsequent division of positive and negative samples for model optimization.

[0060] Based on this, for any target prediction data, the target sample data and target prediction data may be inputted into a preset text annotation model. The text annotation model may be used to process the target prediction data in at least two annotation dimensions, so that the annotation text and prediction scores corresponding to at least two annotation dimensions may be determined based on the processing results. Afterwards, based on the annotation text and the prediction scores corresponding to at least two annotation dimensions, text annotation information corresponding to the target prediction data may be generated to reflect the availability of the target prediction data for subsequent model optimization.

[0061] Following the previous example, after processing the target sample data 1 and the target sample data 2 by multiple language models, the target prediction data 1a and the target prediction data 1b corresponding to the target sample data 1, as well as the target prediction data 2a and the target prediction data 2b corresponding to target sample data 2 are obtained.

5 Based on this, a text annotation model may be used to score the target prediction data in the four dimensions mentioned above and generate detailed text annotations. After scoring and generating annotations for the prediction data corresponding to each sample data by the text annotation model, a prediction score sa1 and an annotation text 1ac corresponding to the target prediction data 1a, a prediction score sb1 and an annotation text 1bc corresponding to the target prediction data 1b, and a prediction score sa2 and an annotation text 2ac
10 corresponding to target prediction data 2a, and a prediction score sb2 and an annotation text 2bc corresponding to target prediction data 2b are obtained. Then, high-quality prediction data and annotation text may be selected based on the above information to construct feedback data for training the language model.

15 **[0062]** In summary, by constructing textual annotation information based on the prediction scores and the annotation texts, the prototype construction of the sample dataset can be realized, which facilitates subsequent model optimization and use.

[0063] In addition, to ensure the quality of the annotation text, quality can be improved by the prediction of candidate text annotation models. In this embodiment, the implementation is
20 specifically described as follows:

determining a preset candidate text annotation model, and processing the target prediction data by the candidate text annotation model, to obtain a candidate annotation text corresponding to the target prediction data; and calculating a text similarity between the candidate annotation text and the annotation text, where in a case that the text similarity is
25 greater than a preset similarity threshold, the step of generating text annotation information corresponding to the target prediction data based on the annotation text and the prediction scores corresponding to the at least two annotation dimensions is performed.

[0064] Specifically, the candidate text annotation model refers to a text annotation model that has the same capability as the text annotation model but has a different model structure
30 from the text annotation model, and is used to generate candidate annotation texts for prediction data, and selecting high-quality annotation texts by comparing the candidate

annotation texts with the annotation texts, for subsequent construction of the sample data array. Correspondingly, the candidate annotation text is an annotation text outputted by the candidate text annotation model. Correspondingly, the text similarity specifically refers to a score used to calculate the similarity between the candidate annotation texts and the annotation texts, which may be obtained by calculating a text distance and a cosine similarity and is not limited in this embodiment.

[0065] Based on this, in order to obtain high-quality annotation texts, the target prediction data may be processed by the candidate text annotation model. In this case, the preset candidate text annotation model may be determined first, and the candidate text annotation model may be used to process the target prediction data to obtain the candidate annotation text corresponding to the target prediction data. On this basis, the text similarity between the candidate annotation text and the annotation text may be calculated. In a case that the text similarity is greater than a preset similarity threshold, it is indicated that the annotation texts outputted by two different text annotation models are highly similar, which can further indicate that the annotation texts are high-quality texts. Therefore, the step of generating text annotation information corresponding to the target prediction data based on the annotation texts and the prediction scores corresponding to the at least two annotation dimensions may be performed.

[0066] In addition, in a case that the text similarity is less than or equal to the preset similarity threshold, it is indicated that the similarity of the annotation texts outputted by the two different text annotation models is small, which in turn indicates that the annotation text outputted by one of the models is not accurate enough. In order to obtain an accurate annotation text, the two models may be used to generate annotation texts separately until the similarity of the obtained annotation texts is greater than the threshold, and then subsequent processing is performed. In addition, if annotation texts with high similarity cannot be obtained for a long time period, a third text annotation model may be introduced for decision-making to ensure the accuracy of the annotation texts, improving the quality of the text annotation information.

[0067] In summary, by constructing the annotation texts based on the candidate text annotation models, high-quality annotation text can be selected, effectively improving the quality of the sample data array, thus achieving better model optimization.

[0068] In step S208, a sample data array is constructed based on the target sample data, the target prediction data and the text annotation information, where the sample data array is used to perform a model training task and a model validation task.

[0069] Specifically, after obtaining the target sample data, the target prediction data outputted by each language model and the text annotation information corresponding to each target prediction data, a sample data array may be constructed based on the target sample data, the target prediction data and the text annotation information, to perform subsequent model training task and model validation task based on the sample data array.

[0070] The sample data array refers to feedback data used for optimizing a language model. Correspondingly, the model training task specifically refers to a task of training a model using a sample data array, and the model validation task specifically refers to a task of validating a model using a sample data array.

[0071] Furthermore, after obtaining the sample data array, in order to improve the quality, data cleaning and format conversion may further be performed, so that the sample data array may be directly reused during an application phase. In this embodiment, the implementation is specifically described as follows:

adding the sample data array to an intermediate sample dataset; performing data cleaning and format conversion on the intermediate sample dataset in a case that the intermediate sample dataset meets a dataset construction condition; and generating a target sample dataset based on a data cleaning result and a format conversion result, where the target sample data array in the target sample dataset is used to perform the model training task and the model validation task.

[0072] Specifically, the intermediate sample dataset refers to a dataset obtained by constructing the sample data array. Correspondingly, the data cleaning specifically refers to a processing operation of removing duplicate and low-quality sample data arrays from the intermediate sample dataset. Correspondingly, the format conversion specifically refers to a processing operation of updating the sample data arrays in the intermediate sample dataset into a unified data structure.

[0073] Based on this, in order to perform model training and model validation by using high-quality data, the sample data array may be added to the intermediate sample dataset after

being obtained. In a case that the intermediate sample dataset meets a dataset construction condition, it is indicated that the sample data array in the intermediate sample dataset may be used for subsequent tasks. In this case, data cleaning and format conversion may be performed on the intermediate sample dataset first, to remove redundant data and unifying the data format, so that the target sample dataset can be generated based on a data cleaning result and a format conversion result. Subsequently, the model training task and the model validation task may be performed based on the target sample data array in the target sample dataset.

[0074] In summary, by performing data cleaning and format conversion on the intermediate sample dataset, a pure and high-quality sample data array may be obtained with the same data format, which is more convenient for subsequent model training and evaluation.

[0075] In addition, after obtaining a high-quality target sample data array, in order to simulate real model optimization scenarios, a reward model and a criticism model may be trained by using data first, and the language model may be optimized based on the trained model. In this embodiment, the implementation is specifically described as follows:

determining a reward model and a criticism model corresponding to a language model to be optimized; training the reward model and the criticism model based on the target sample array in the target sample dataset, and determining a target reward model and a target criticism model based on a training result; and optimizing the language model to be optimized by using the target reward model and the target criticism model until obtaining a target language model that meets an optimization stopping condition.

[0076] Specifically, the reward model and the criticism model refer to auxiliary models used to optimize the language model to be optimized. The reward model is used to evaluate and improve the prediction quality of the language model to be optimized, and the criticism model is used to provide detailed improvement suggestions, thereby enabling the language model to achieve self-learning capability. Correspondingly, the optimization stopping condition specifically refers to a condition for stopping the optimization of the language model to be optimized, such as the number of iterations, loss value comparison condition, or validation set validation condition, which is not limited in this embodiment.

[0077] Based on this, in order to effectively optimize the language model to be optimized, the reward model and the criticism model corresponding to the language model to be

optimized may be determined first. In this case, the reward model and the criticism model may be trained based on the target sample data array in the target sample dataset, and the target reward model and the target criticism model may be determined based on the training results. On this basis, the language model to be optimized may be optimized by the target
5 reward model and the target criticism model, to adjust the model to be optimized to the target language model that meets the optimization stopping condition, which facilitates the use in practical business scenarios.

[0078] Following the previous example, after obtaining the prediction data, the prediction scores and the annotation texts corresponding to the target sample data 1 and the target sample
10 data 2 respectively, an intermediate sample dataset may be constructed based on various dimensions of data. After performing data cleaning and format conversion on the dataset, a high-quality target sample dataset may be obtained. On this basis, the reward model and the criticism model of the LLM may be trained by using the target sample dataset, so that the reward model can improve the output quality of the LLM model, and the criticism model can
15 provide detailed model improvement suggestions. Based on this, the LLM model is optimized, so that the target LLM model that meets the usage condition can be obtained. In the application stage, if a user inputs a question text "Where is the capital of a country A", the model may provide an answer text "city B".

[0079] With the data processing method according to the embodiment, in order to provide
20 abundant and diversified feedback data for model optimization of language models, the feedback data may be constructed by using a text annotation model and multiple language models. Specifically, sampling may be performed for multiple datasets, to determine target sample data based on a sampling result, the sample diversity can be improved by using multiple datasets. Thereafter, the target sample data may be inputted to the multiple language
25 models for processing, and the target prediction data outputted from each of the multiple language models may be obtained. On this basis, text annotation processing may be performed on the target prediction data by a preset text annotation model, to obtain the text annotation information corresponding to the target prediction data. In this way, when constructing the text annotation information, in order to ensure the quality and scale of the
30 data, multiple language models may be used to output multiple pieces of target prediction information, to effectively ensure the abundance of the data. Then, a sample data array may be constructed based on the target sample data, the target prediction data and the text annotation

information corresponding to the target prediction data, so that the model training task and the model validation task can be performed in the application stage based on the sample data array, thereby improving the performance and alignment effect of the language model.

[0080] Refer to Figure 3, Figure 3 shows a flowchart of a model optimization method according to an embodiment of the present disclosure, which includes the following steps S302 to S306.

[0081] In step S302, a reward model and a criticism model corresponding to a language model to be optimized is determined.

[0082] In step S304, the reward model and the criticism model is trained based on the target sample data array in the target sample dataset, and a target reward and a target criticism model are determined based on training results. The target sample data array in the target sample dataset is obtained through the above method.

[0083] In step S306, the language model to be optimized is optimized by the target reward model and the target criticism model until a target language model that meets an optimization stopping condition is obtained.

[0084] For description of the model optimization method in this embodiment, reference may be made to the description of the data processing method in the above embodiment, which is not repeated here.

[0085] The data processing method is further described by taking the application of the data processing method according to the present disclosure in scenarios of large language models as an example with reference to Figure 4. Figure 4 shows a processing procedure of a data processing method according to an embodiment of the present disclosure, which specifically includes the following steps S402 to S422.

[0086] In step S402, multiple datasets are determined, where data types of sample data in the multiple datasets are different from each other.

[0087] In step S404, sampling is performed for each dataset to obtain multiple pieces of initial sample data, and an initial sample dataset is constructed based on the multiple pieces of initial sample data.

[0088] In step S406, target sample data is selected from the initial sample dataset.

[0089] In step S408, the target sample data is inputted into multiple language models for processing, to obtain the target prediction data outputted by each language model.

[0090] In step S410, text annotation processing is performed on each target prediction data by a preset text annotation model to obtain text annotation information corresponding to each target prediction data.

[0091] The text annotation information corresponding to the target prediction data outputted by each of the language models is determined by: inputting the target sample data and the target prediction data into a preset text annotation model, processing the target prediction data in at least two annotation dimensions by the text annotation model, and determining, based on a processing result, an annotation text and prediction scores corresponding to the at least two annotation dimensions; and generating text annotation information corresponding to the target prediction data based on the annotation text and the prediction scores corresponding to the at least two annotation dimensions.

[0092] In step S412, a sample data array is constructed based on the target sample data, target prediction data and text annotation information, and the sample data array is added to the intermediate sample dataset.

[0093] In step S414, if the intermediate sample dataset meets a dataset construction condition, data cleaning and format conversion are performed on the intermediate sample dataset.

[0094] In step S416, a target sample dataset is generated based on a data cleaning result and a format conversion result.

[0095] In step S418, in a case that a model optimization task is triggered, a reward model and a criticism model corresponding to a language model to be optimized is determined.

[0096] In step S420, based on a target sample data array in the target sample dataset, the reward model and the criticism model are trained, and a target reward and a target criticism model are determined based on training results.

[0097] In step S422, the language model to be optimized is optimized by using the target reward model and the target criticism model until obtaining a target language model that meets a optimization stopping condition.

[0098] With the data processing method according to the embodiment, in order to provide

abundant and diversified feedback data for model optimization of language models, the feedback data may be constructed by using a text annotation model and multiple language models. Specifically, sampling may be performed for multiple datasets, to determine target sample data based on a sampling result, the sample diversity can be improved by using multiple datasets. Thereafter, the target sample data may be inputted to the multiple language models for processing, and the target prediction data outputted from each of the multiple language models may be obtained. On this basis, text annotation processing may be performed on the target prediction data by a preset text annotation model, to obtain the text annotation information corresponding to the target prediction data. In this way, when constructing the text annotation information, in order to ensure the quality and scale of the data, multiple language models may be used to output multiple pieces of target prediction information, to effectively ensure the abundance of the data. Then, a sample data array may be constructed based on the target sample data, the target prediction data and the text annotation information corresponding to the target prediction data, so that the model training task and the model validation task can be performed in the application stage based on the sample data array, thereby improving the performance and alignment effect of the language model.

[0099] Corresponding to the above method embodiments, an embodiment of a data processing device is further provided according to the present disclosure. Figure 5 shows a schematic diagram of a structure of a data processing device according to an embodiment of the present disclosure. As shown in Figure 5, the device includes: a determination module 502, a processing module 504, an annotation module 506 and a construction module 508. The determination module 502 is configured to perform sampling for multiple datasets, and determine target sample data based on a sampling result. The processing module 504 is configured to input the target sample data into multiple language models for processing, to obtain target prediction data outputted by each of the language models. The annotation module 506 is configured to perform text annotation processing on the target prediction data by a preset text annotation model, to obtain text annotation information corresponding to the target prediction data. The construction module 508 is configured to construct a sample data array based on the target sample data, the target prediction data and the text annotation information, where the sample data array is used to perform a model training task and a model validation task.

[0100] In a preferred embodiment, the determination module 502 is further configured to:

determine the multiple datasets, where each of the datasets includes sample data of a different data type from others; perform sampling for each of the datasets to obtain multiple pieces of initial sample data, and construct an initial sample dataset based on the multiple pieces of initial sample data; and select the target sample data from the initial sample dataset.

5 **[0101]** In a preferred embodiment, the device further includes:

an information determination module, configured to select the multiple language models for the target sample data and determine prompt information corresponding to each of the language models, where processing the target sample data by any one of the multiple language models includes: updating the target sample data based on the prompt information
10 corresponding to the language model, and generating model input context information based on an update result; and inputting the model input context information into the language model for processing, to obtain the target prediction data.

[0102] In a preferred embodiment, the annotation module 506 is further configured to:

input the target sample data and the target prediction data into a preset text annotation
15 model, process the target prediction data in at least two annotation dimensions by the text annotation model, and determine, based on a processing result, an annotation text and prediction scores corresponding to the at least two annotation dimensions; and generate text annotation information corresponding to the target prediction data based on the annotation text and the prediction scores corresponding to the at least two annotation dimensions.

20 **[0103]** In a preferred embodiment, the device further includes:

a similarity calculation module, configured to determining a preset candidate text annotation model, and processing the target prediction data by the candidate text annotation model, to obtain a candidate annotation text corresponding to the target prediction data; and calculating a text similarity between the candidate annotation text and the annotation text,
25 where in a case that the text similarity is greater than a preset similarity threshold, the step of generating text annotation information corresponding to the target prediction data based on the annotation text and the prediction scores corresponding to the at least two annotation dimensions is performed.

[0104] In a preferred embodiment, the device further includes:

30 a dataset construction module, configured to add the sample data array to an intermediate

sample dataset;

perform data cleaning and format conversion on the intermediate sample dataset in a case that the intermediate sample dataset meets a dataset construction condition; and

generate a target sample dataset based on a data cleaning result and a format conversion result, where the target sample data array in the target sample dataset is used to perform the model training task and the model validation task.

[0105] In a preferred embodiment, the device further includes:

a model optimization module, configured to determine a reward model and a criticism model corresponding to a language model to be optimized; train the reward model and the criticism model based on the target sample array in the target sample dataset, and determine a target reward model and a target criticism model based on a training result; and optimize the language model to be optimized by using the target reward model and the target criticism model until a target language model that meets an optimization stopping condition is obtained.

[0106] With the data processing device according to the embodiment, in order to provide abundant and diversified feedback data for model optimization of language models, the feedback data may be constructed by using a text annotation model and multiple language models. Specifically, sampling may be performed for multiple datasets, to determine target sample data based on a sampling result, the sample diversity can be improved by using multiple datasets. Thereafter, the target sample data may be inputted to the multiple language models for processing, and the target prediction data outputted from each of the multiple language models may be obtained. On this basis, text annotation processing may be performed on the target prediction data by a preset text annotation model, to obtain the text annotation information corresponding to the target prediction data. In this way, when constructing the text annotation information, in order to ensure the quality and scale of the data, multiple language models may be used to output multiple pieces of target prediction information, to effectively ensure the abundance of the data. Then, a sample data array may be constructed based on the target sample data, the target prediction data and the text annotation information corresponding to the target prediction data, so that the model training task and the model validation task can be performed in the application stage based on the sample data array, thereby improving the performance and alignment effect of the language model.

[0107] The above is a schematic solution of a data processing device in this embodiment. It should be noted that the technical solution of the data processing device belongs to the same concept as the technical solution of the data processing method mentioned above. For details not described in detail in the technical solution of the data processing device, please refer to the description of the technical solution of the data processing method mentioned above.

[0108] Corresponding to the above method embodiments, an embodiment of a model optimization device is further provided according to the present disclosure. Figure 6 is a schematic diagram of a structure of a model optimization device according to an embodiment of the present disclosure. As shown in Figure 6, the device includes: a model determination module 602, a model training module 604 and a model optimization module 606. The model determination module 602 is configured to determine a reward model and a criticism model corresponding to a language model to be optimized. The model training module 604 is configured to train the reward model and the criticism model based on the target sample array in the target sample dataset, and determine a target reward model and a target criticism model based on a training result, where the target sample data array in the target sample dataset is obtained with the above method. The model optimization module 606 is configured to optimize the language model to be optimized by using the target reward model and the target criticism model until a target language model that meets an optimization stopping condition is obtained.

[0109] A schematic solution of the model optimization device according to the embodiment is described above. It should be noted that the technical solution of the model optimization device and the technical solution of the model optimization method described above belong to the same concept, and the details not described in detail in the technical solution of the model optimization device may be referred to the description of the technical solution of the model optimization method described above.

[0110] Figure 7 illustrates a block diagram of a structure of a computing device 700 according to an embodiment of the present disclosure. The components of the computing device 700 include, but are not limited to, a memory 710 and a processor 720. The processor 720 is connected to the memory 710 via a bus 730, and a database 750 is used to store data.

[0111] The computing device 700 further includes an access device 740, the access device 740 enables the computing device 700 to communicate via one or more networks 760.

Examples of these networks include a public switched telephone network (PSTN), a local area network (LAN), a wide area network (WAN), a personal area network (PAN), or a combination of communication networks such as the Internet. The access device 740 may include one or more of any types of wired or wireless network interfaces (e.g., a network interface controller (NIC), such as an IEEE802.11, Wireless Local Area Network (WLAN) wireless interface, Worldwide Interoperability for Microwave Access (Wi-MAX) interface, Ethernet interface, Universal Serial Bus (USB) interface, cellular network interface, Bluetooth interface, Near Field Communication (NFC).

[0112] In an embodiment of the present disclosure, the above components of the computing device 700, as well as other components not shown in Figure 7, may also be connected to each other, such as via a bus. It should be understood that the block diagram of the structure of the computing device shown in Figure 7 is for exemplary purposes only and is not a limitation on the scope of the present disclosure. Those skilled in the art may add or replace other components as desired.

[0113] The computing device 700 may be any type of stationary or mobile computing device, including a mobile computer or a mobile computing device (e.g., a tablet computer, a personal digital assistant, a laptop computer, a notebook computer, a netbook, etc.), a cell phone (e.g., a smartphone), a wearable computing device (e.g., a smartwatch, smartglasses, etc.), or other types of mobile devices, or a stationary computing device such as desktop computer or a personal computer (PC). The computing device 700 may also be a mobile or stationary server.

[0114] The processor 720 is used to execute computer executable instructions, which, when executed by the processor, perform the steps of the data processing method or the model optimization method described above.

[0115] A schematic solution of the computing device of the other embodiment is described above. It should be noted that the technical solution of the computing device and the technical solution of the data processing method or the model optimization method described above belong to the same concept, and the details not described in detail in the technical solution of the computing device may be referred to the description of the technical solution of the data processing method or the model optimization method described above.

[0116] A computer-readable storage medium is further provided according to an

embodiment of the present disclosure, which stores computer-executable instructions, the computer-executable instructions, when executed by a processor, perform the steps of the data processing method or the model optimization method described above.

5 [0117] A schematic solution of the computer-readable storage medium of the other embodiment is described above. It should be noted that the technical solution of the computer-readable storage medium and the technical solution of the data processing method or the model optimization method described above belong to the same concept, and the details not described in detail in the technical solution of the computer-readable storage medium may be referred to the description of the technical solution of the data processing method or the
10 model optimization method described above.

[0118] A computer program product is further provided according to an embodiment of the present disclosure, which includes computer programs or instructions, the computer programs or instructions, when executed by a processor, perform the steps of the data processing method or the model optimization method described above.

15 [0119] A schematic solution of the computer program product of the other embodiment is described above. It should be noted that the technical solution of the computer program product and the technical solution of the data processing method or the model optimization method described above belong to the same concept, and the details not described in detail in the technical solution of computer program product may be referred to the description of the
20 technical solution of the data processing method or the model optimization method described above.

[0120] Preferred embodiments of the present disclosure are described above. Other embodiments are within the scope of the claims. In some cases, the actions or steps described in the claims may be performed in a different order than in the embodiments and still achieve
25 the desired results. In addition, the processes depicted in the drawings do not necessarily require the particular order or consecutive order shown to achieve the desired results. In some embodiments, multi-tasking and parallel processing is also possible or may be advantageous.

[0121] The computer instruction includes computer program codes, the computer program code may be in the form of source code, in the form of object code, in the form of an
30 executable file or in some intermediate form or the like. The computer readable medium may include: any entity or device capable of carrying the computer program code, a recording

medium, a USB disk, a removable hard disk, a diskette, an optical disk, a computer memory, a Read-Only Memory (ROM), a Random Access Memory (RAM), an electric carrier signal, a telecommunications signal, and software distribution medium, etc. It should be noted that the content of the computer-readable medium may be increased or decreased according to the requirements of the patent practice, for example, in some regions, according to the patent practice, the computer-readable medium does not include carrier signals and telecommunication signals.

[0122] It should be noted that, for each of the foregoing method embodiments, for the sake of simplicity of description, they are all expressed as a series of combinations of actions, but those skilled in the art should know that the embodiments of the present disclosure are not limited by the order of the described actions, because according to the embodiments of the present disclosure, certain steps may be performed in other orders or simultaneously. Secondly, those skilled in the art should also be aware that the embodiments described in the disclosure are preferred embodiments, and the actions and modules involved are not necessarily necessary for the embodiments of the present disclosure.

[0123] In the above-described embodiments, the description of each embodiment has its own focus, and the part not described in detail in a certain embodiment can be referred to the relevant description of other embodiments.

[0124] The above preferred embodiments of the present disclosure are intended only to aid in the description of the present disclosure. The preferred embodiments are not an exhaustive recitation of all details, nor do they limit the invention to the specific embodiments described only. Apparently, many modifications and variations may be made in accordance with the contents of the embodiments of the present disclosure. These embodiments are selected and specifically described in the present disclosure to better explain the principles and practice of the embodiments of the present disclosure so that those skilled in the art to which they belong will be able to well understand and utilize the present disclosure.

CLAIMS

1. A data processing method, comprising:

performing sampling for a plurality of datasets, and determining target sample data based

5 on a sampling result;

inputting the target sample data into a plurality of language models for processing, to obtain target prediction data outputted by each of the language models;

performing text annotation processing on the target prediction data by a preset text annotation model, to obtain text annotation information corresponding to the target prediction

10 data; and

constructing a sample data array based on the target sample data, the target prediction data and the text annotation information, wherein the sample data array is used to perform a model training task and a model validation task.

2. The data processing method according to claim 1, wherein the performing sampling

15 for a plurality of datasets, and determining target sample data based on a sampling result comprises:

determining the plurality of datasets, wherein each of the datasets comprises sample data of a different data type from other datasets;

20 performing sampling for each of the datasets to obtain a plurality of pieces of initial sample data, and constructing an initial sample dataset based on the plurality of pieces of initial sample data; and

selecting the target sample data from the initial sample dataset.

3. The data processing method according to claim 1, wherein before the inputting the target sample data into a plurality of language models for processing, to obtain target

25 prediction data outputted by each of the language models, the method further comprises:

selecting the plurality of language models for the target sample data and determining prompt information corresponding to each of the language models, wherein

processing the target sample data by any one of the plurality of language models comprises:

updating the target sample data based on the prompt information corresponding to the language model, and generating model input context information based on an update result; and

5 inputting the model input context information into the language model for processing, to obtain the target prediction data.

4. The data processing method according to claim 1, wherein the text annotation information corresponding to the target prediction data outputted by each of the language models is determined by:

10 inputting the target sample data and the target prediction data into a preset text annotation model, processing the target prediction data in at least two annotation dimensions by the text annotation model, and determining, based on a processing result, an annotation text and prediction scores corresponding to the at least two annotation dimensions; and

15 generating the text annotation information corresponding to the target prediction data based on the annotation text and the prediction scores corresponding to the at least two annotation dimensions.

5. The data processing method according to claim 4, further comprising:

 determining a preset candidate text annotation model, and processing the target prediction data by the candidate text annotation model, to obtain a candidate annotation text corresponding to the target prediction data; and

20 calculating a text similarity between the candidate annotation text and the annotation text, wherein

 in a case that the text similarity is greater than a preset similarity threshold, the step of generating the text annotation information corresponding to the target prediction data based on the annotation text and the prediction scores corresponding to the at least two annotation dimensions is performed.

25 6. The data processing method according to claim 1, wherein after the constructing a sample data array based on the target sample data, the target prediction data and the text annotation information, the method further comprises:

 adding the sample data array to an intermediate sample dataset;

performing data cleaning and format conversion on the intermediate sample dataset in a case that the intermediate sample dataset meets a dataset construction condition; and

generating a target sample dataset based on a data cleaning result and a format conversion result, wherein the target sample data array in the target sample dataset is used to perform the model training task and the model validation task.

7. The data processing method according to claim 6, further comprising:

determining a reward model and a criticism model corresponding to a language model to be optimized;

training the reward model and the criticism model based on the target sample array in the target sample dataset, and determining a target reward model and a target criticism model based on a training result; and

optimizing the language model to be optimized by using the target reward model and the target criticism model until obtaining a target language model that meets an optimization stopping condition.

8. A model optimization method, comprising:

determining a reward model and a criticism model corresponding to a language model to be optimized;

training the reward model and the criticism model based on a target sample array in a target sample dataset, and determining a target reward model and a target criticism model based on a training result, wherein the target sample data array in the target sample dataset is obtained with the method according to any one of claims 1 to 7; and

optimizing the language model to be optimized by using the target reward model and the target criticism model until obtaining a target language model that meets an optimization stopping condition.

9. A data processing device, comprising:

a determination module, configured to perform sampling for a plurality of datasets, and determine target sample data based on a sampling result;

a processing module, configured to input the target sample data into a plurality of language models for processing, to obtain target prediction data outputted by each of the

language models;

an annotation module, configured to perform text annotation processing on the target prediction data by a preset text annotation model, to obtain text annotation information corresponding to the target prediction data; and

5 a construction module, configured to construct a sample data array based on the target sample data, the target prediction data and the text annotation information, wherein the sample data array is used to perform a model training task and a model validation task.

10. A model optimization device, comprising:

a model determination module, configured to determine a reward model and a criticism
10 model corresponding to a language model to be optimized;

a model training module, configured to train the reward model and the criticism model based on the target sample array in the target sample dataset, and determine a target reward model and a target criticism model based on a training result, wherein the target sample data array in the target sample dataset is obtained with the method according to any one of claims
15 1 to 7; and

a model optimization module, configured to optimize the language model to be optimized by using the target reward model and the target criticism model until a target language model that meets an optimization stopping condition is obtained.

11. A computing device, comprising:

20 a memory, configured to store computer executable instructions; and

a processor, configured to execute the computer executable instructions, wherein the computer executable instructions, when executed by the processor, perform the steps of the method according to any one of claims 1 to 8.

12. A computer-readable storage medium, comprising computer executable instructions
25 stored thereon, wherein the computer executable instructions, when executed by a processor, perform the steps of the method according to any one of claims 1 to 8.

13. A computer program product, comprising computer programs or instructions, wherein the computer programs or instructions, when executed by a processor, perform the steps of the method according to any one of claims 1 to 8.