

DATA PROCESSING METHOD AND DEVICE

[0001] The present application claims priority to Chinese Patent Application No. 202411191435.9, titled "DATA PROCESSING METHOD AND DEVICE", filed on August 28, 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.

BACKGROUND

[0003] With the development of computer and Internet technologies, the deep learning technology has been applied in more and more scenarios. Deep learning typically focuses on digital media files that are easy to be understood by humans, such as texts, audios, and images. Most existing deep learning models are operated in the field of media data, for example, the language model performs well in text generation and understanding through token prediction, the audio processing model achieves advanced speech recognition and speech to speech translation by combining texts and speeches, and the image processing model generates images by predicting next pixels. These models have achieved significant results in their respective fields, but the ubiquitous native binary data in the digital world is precisely ignored. As a basis of all digital data, devices and software, processing from computer processors to operating systems in common electronic products are all performed based on bytes. However, the most essential elements of data is ignored in the model prediction process in the conventional technology, resulting in a lack of comprehensive development of prediction capability and speed of the model. Therefore, an effective solution is urgently needed to solve 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 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.

5 **[0005]** In a first aspect of the embodiments according to the present disclosure, a data processing method is provided, which includes:

 acquiring multimedia data of a target prediction task, and converting the multimedia data into a binary byte sequence;

 inputting the binary byte sequence into a multimedia processing model, where the
10 multimedia processing model includes a linear projection unit, a sequence decoding unit and a byte decoding unit;

 converting, by the linear projection unit, the binary byte sequence into a byte sequence vector, and processing the byte sequence vector by the sequence decoding unit to obtain a sequence feature; and

15 processing the sequence feature by the byte decoding unit to obtain prediction bytes, and constructing target task data corresponding to the target prediction task based on the prediction bytes.

[0006] Preferably, the converting the multimedia data into a binary byte sequence includes:

 determining code point information corresponding to the multimedia data based on a
20 conversion strategy associated with the target prediction task; and

 processing the multimedia data based on the code point information and an encoding rule configured in the conversion strategy, to obtain the binary byte sequence corresponding to the multimedia data.

[0007] Preferably, the converting, by the linear projection unit, the binary byte sequence
25 into a byte sequence vector includes:

 inputting the binary byte sequence into the linear projection unit, and converting, by the linear projection unit, binary bytes in the binary byte sequence into byte vectors;

 constructing an initial byte sequence vector corresponding to the binary byte sequence based on a byte vector corresponding to each of the binary bytes; and

converting the initial byte sequence vector into a one-dimensional vector, and mapping the one-dimensional vector to a binary byte space to obtain the byte sequence vector.

[0008] Preferably, the processing the byte sequence vector by the sequence decoding unit to obtain a sequence feature includes:

5 inputting the byte sequence vector into the sequence decoding unit, and performing autoregressive prediction on the byte sequence vector by the sequence decoding unit to obtain the sequence feature,

 where the processing the sequence feature by the byte decoding unit to obtain prediction bytes includes:

10 inputting the sequence feature into the byte decoding unit, and performing autoregressive reconstruction on the sequence feature by the byte decoding unit to obtain the prediction bytes.

[0009] Preferably, the constructing target task data corresponding to the target prediction task based on the prediction bytes includes:

15 determining an inverse conversion strategy corresponding to the conversion strategy, and parsing the prediction bytes based on the inverse conversion strategy; and

 determining prediction code point information corresponding to the prediction bytes based on a parsing result, and generating the target task data corresponding to the target prediction task based on the prediction code point information.

20 **[0010]** Preferably, before the inputting the binary byte sequence into a multimedia processing model, the method further includes:

 acquiring sample data and a sample label corresponding to the sample data, converting the sample data into a sample binary byte sequence, and converting the sample label into label bytes;

25 inputting the sample binary byte sequence into an initial multimedia processing model for processing, to obtain training prediction bytes; and

 training the initial multimedia processing model based on the training prediction bytes and the label bytes until obtaining the multimedia processing model that meets the training stopping condition.

[0011] In a second aspect of the embodiments according to the present disclosure, a data processing device is provided, which includes: an acquisition module, an input module, a conversion module and a construction module. The acquisition module is configured to acquire multimedia data of a target prediction task, and convert the multimedia data into a binary byte sequence. The input module is configured to input the binary byte sequence into a multimedia processing model, where the multimedia processing model includes a linear projection unit, a sequence decoding unit and a byte decoding unit. The conversion module is configured to convert, by the linear projection unit, the binary byte sequence into a byte sequence vector, and process the byte sequence vector by the sequence decoding unit to obtain a sequence feature. The construction module is configured to process the sequence feature by the byte decoding unit to obtain prediction bytes, and construct target task data corresponding to the target prediction task based on the prediction bytes.

[0012] In a third 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 described above.

[0013] In a fourth 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 described above.

[0014] In a fifth 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 described above.

[0015] With the data processing method according to the embodiment, in order to improve the comprehensive understanding and simulation of digital systems by the model, so as to enhance the processing capability and efficiency of multimedia data, after obtaining the multimedia data of the target prediction task, the multimedia data may be first converted into a binary byte sequence. In this case, the binary byte sequence may be inputted into the multimedia processing model, which includes a linear projection unit, a sequence decoding

unit and a byte decoding unit. On this basis, the binary byte sequence may be converted into a byte sequence vector by the linear projection unit, and the byte sequence vector is processed by the sequence decoding unit to obtain a sequence feature, thereby exploring a structural relationship at a byte sequence dimension. The sequence feature is processed by the byte decoding unit, to obtain prediction bytes, thereby exploring a byte relationship at a byte dimension. Finally, target task data corresponding to the target prediction task may be constructed based on the prediction bytes, achieving processing at a binary byte dimension by the model. Therefore, the integration process between different types of data is simplified, improving the applicability and performance of the model in cross modal tasks.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] Figure 1 is a schematic diagram of a data processing method according to an embodiment of the present disclosure;

[0017] Figure 2 is a flowchart of a data processing method according to an embodiment of the present disclosure;

[0018] Figure 3 is a flowchart of a processing procedure of a data processing method according to an embodiment of the present disclosure;

[0019] Figure 4 is a schematic diagram of a structure of a data processing device according to an embodiment of the present disclosure; and

[0020] Figure 5 is a block diagram of a structure of a computing device according to an embodiment of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

[0021] 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.

[0022] 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 refers to and encompasses any or all possible combinations of one or more of the listed items in association.

[0023] 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 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".

[0024] 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.

[0025] In the present disclosure, a data processing method is provided. The present disclosure further relates to a data processing device, a computing device, a computer-readable storage medium and a computer program product, which will be described in detail in the following embodiments.

[0026] In practice, deep learning models mainly focus on processing digital media data that is easy to be understood by humans, such as texts, audios, and images. Although these models have achieved significant results in their respective fields, while have obvious shortcomings in processing native binary data. Especially, in the text processing aspect, the language model performs well in text generation and understanding through token prediction, but relies on the token process of texts and cannot directly process binary data. In the audio processing aspect,

although the audio processing model may process audio data, the audio processing model is also limited to specific data types and cannot effectively process and simulate binary data. In the image processing aspect, although the image processing model may predict the next pixel in an image, the image processing model performs poorly when dealing with long byte sequences and complex binary data. Therefore, an effective solution is urgently needed to solve the above problem.

[0027] Referring to the schematic diagram shown in Figure 1, with the data processing method according to the embodiment, in order to improve the comprehensive understanding and simulation of digital systems by the model, so as to enhance the processing capability and efficiency of multimedia data, after obtaining the multimedia data of the target prediction task, the multimedia data may be first converted into a binary byte sequence. In this case, the binary byte sequence may be inputted into the multimedia processing model, which includes a linear projection unit, a sequence decoding unit and a byte decoding unit. On this basis, the binary byte sequence may be converted into a byte sequence vector by the linear projection unit, and the byte sequence vector is processed by the sequence decoding unit to obtain a sequence feature, thereby exploring a structural relationship at a byte sequence dimension. The sequence feature is processed by the byte decoding unit, to obtain prediction bytes, thereby exploring a byte relationship at a byte dimension. Finally, target task data corresponding to the target prediction task may be constructed based on the prediction bytes, achieving processing at a binary byte dimension by the model. Therefore, the integration process between different types of data is simplified, improving the applicability and performance of the model in cross modal tasks.

[0028] Referring to Figure 2, Figure 2 shows a flowchart of a data processing method according to an embodiment of the present disclosure, which includes the following steps S202 to S208.

[0029] In step S202, multimedia data of a target prediction task is acquired, and the multimedia data is converted into a binary byte sequence.

[0030] The data processing method according to this embodiment may be applied to model applications in any scenarios, such as text processing scenarios (abstract extraction, text classification, translation), image processing scenarios (image cropping, image generation, graphic and text generation), audio processing scenarios (noise elimination, audio generation,

audio conversion), which is not limited in this embodiment. The data processing method is described by taking an application of data processing method in the text processing scenario as an example in this embodiment. For description of other scenarios, reference may be made to the same or corresponding content in this embodiment, and is not limited in this
5 embodiment. In addition, the data processing method according to this embodiment may also be applied to cross modal scenarios.

[0031] Specifically, the target prediction task refers to a task triggered in a multimedia data processing scenario, which may be a text processing task in a text processing scenario such as abstract extraction, text classification, and translation. The target prediction task may also be
10 an image processing task in an image processing scenario, such as image cropping, image generation, graphic and text generation. The target prediction task may also be an audio processing task in an audio processing scenario, such as noise reduction, audio generation, and audio conversion. Correspondingly, the multimedia data refers to data that needs to be inputted into the model for processing when triggering the target prediction task, such as texts,
15 images, videos, and audio. Correspondingly, the binary byte sequence specifically refers to a sequence of binary bytes obtained by encoding the multimedia data into bytes, which is the most basic unit of computer processing. By converting the multimedia data into binary bytes for model processing, prediction efficiency may be effectively improved. In addition, the model may be effectively developed, achieving multimedia data analysis from the most basic
20 unit.

[0032] Based on this, in order to improve the comprehensive understanding and simulation of digital systems by the model, to enhance the processing capability and efficiency of multimedia data, after obtaining the multimedia data of the target prediction task, the multimedia data may be first converted into a binary byte sequence. In this case, the binary
25 byte sequence may be inputted into the multimedia processing model, which includes a linear projection unit, a sequence decoding unit and a byte decoding unit. On this basis, the binary byte sequence may be converted into byte sequence vectors by the linear projection unit, and the byte sequence vectors are processed by the sequence decoding unit to obtain a sequence feature, thereby determining the structural relationship in a byte sequence dimension. The
30 sequence feature is processed by the byte decoding unit, to obtain prediction bytes, thereby determining the byte relationship in a byte dimension. Finally, target task data corresponding to the target prediction task may be constructed based on the prediction bytes, achieving

binary byte processing through the model.

[0033] Furthermore, when constructing the binary byte sequence for the multimedia data, in order to achieve cross modal and accurate model prediction, a preset conversion strategy may be used. In this embodiment, the implementation is specifically described as follows:

5 determining code point information corresponding to the multimedia data based on a conversion strategy associated with the target prediction task; and processing the multimedia data based on the code point information and an encoding rule configured in the conversion strategy, to obtain the binary byte sequence corresponding to the multimedia data.

[0034] Specifically, the conversion strategy specifically refers to a strategy used to convert
10 the multimedia data associated with the target prediction task into a binary byte sequence in a specified encoding manner. This strategy may be used to determine the code point information for each character in the multimedia data and the encoding rule, where the encoding rule is used to encode the multimedia data to obtain binary bytes to form the binary byte sequence.

15 **[0035]** Based on this, after obtaining the multimedia data of the target prediction task, the code point information corresponding to the multimedia data may be determined according to the conversion strategy associated with the target prediction task. On this basis, the multimedia data may be processed based on the code point information and the encoding rule configured in the conversion strategy, to convert the multimedia data into multiple binary
20 bytes, thereby forming a binary byte sequence corresponding to the multimedia data for subsequent input to the mode for prediction processing.

[0036] For example, a user provides a question text "Who is the author of the ancient poem A" through a client, upon receiving the question text, the code point corresponding each character unit in Unicode encoding may be determined first. On this basis, 8-bit Unicode
25 Transformation Format (UTF-8) may be used to encode each character unit. In UTF-8 encoding, the code point corresponding to each character unit is converted into a sequence of bytes of a set length. By calculating the byte sequence encoded in UTF-8 encoding and performing byte allocation, a hexadecimal byte sequence corresponding to the question text
30 "Who is the author of the ancient poem A" may be obtained, which is expressed as {A\xe5\x8f\xa4\xe8\xaf\x97\xe7\x9a\x84\xe4\xbd\x9c\xe8\x80\x85\xe6\x98\xaf\xe8\xb0\x81}. Based on this, hexadecimal byte sequence may be converted into a binary byte sequence for

subsequent model processing.

[0037] In summary, the multimedia data is converted by using a conversion strategy, so that the multimedia data can be converted into a unified format binary byte sequence according to a set specification, thereby achieving subsequent accurate prediction of the model.

5 **[0038]** In step S204, the binary byte sequence is inputted into a multimedia processing model, where the multimedia processing model includes a linear projection unit, a sequence decoding unit, and a byte decoding unit.

[0039] In step S206, the binary byte sequence is converted into a byte sequence vector by the linear projection unit, and the byte sequence vector is processed by the sequence decoding
10 unit to obtain a sequence feature.

[0040] In step S208, the sequence feature is processed by the byte decoding unit, to obtain prediction bytes, and target task data corresponding to the target prediction task is constructed based on the prediction bytes.

[0041] Specifically, after obtaining the binary byte sequence corresponding to the multimedia data mentioned above, the binary byte sequence may be inputted into the
15 multimedia processing model that includes the linear projection unit, the sequence decoding unit and the byte decoding unit. In this way, the multimedia processing model can output prediction bytes by performing binary byte level prediction to construct the target task data corresponding to the target prediction task for subsequent use.

20 **[0042]** The linear projection unit specifically refers to a unit used to map the binary bytes in the binary byte sequence into a vector expression. Correspondingly, the sequence decoding unit specifically refers to a decoder that constructs a sequence dimensional feature, and the byte decoding unit specifically refers to a decoder that constructs a byte dimensional feature for outputting prediction bytes. Correspondingly, the byte sequence vector specifically refers
25 to a sequence vector obtained by performing vector mapping on the binary byte sequence. The sequence feature specifically refers to a sequence level feature extracted from the byte sequence vector, and the prediction byte specifically refers to a model output result obtained by processing the multimedia data corresponding to the target prediction task, which are binary bytes used to construct the target task data. Correspondingly, the target task data
30 specifically refers to a result obtained after model processing of the multimedia data. For

example, if the target prediction task is a text generation task, the target task data is a text generation result. For example, if the target prediction task is an audio noise reduction task, the target task data is an audio with noise reduction.

[0043] That is, when processing the binary byte sequence by the multimedia processing model, in order to determine the byte relationship at a binary byte level, the binary byte sequence may be first converted into a byte sequence vector by a linear projection unit. In a case that the vector expression corresponding to the binary byte sequence is obtained, subsequent prediction processing may be performed. In order to improve the prediction capability of the model and achieve sufficient data relationship analysis of the binary bytes, the byte sequence vector may be processed first by the sequence decoding unit to obtain the sequence feature of a corresponding sequence dimension. The sequence feature is processed by the byte decoding unit, prediction processing may be performed to obtain the prediction bytes. Afterwards, based on the prediction bytes, the target task data corresponding to the target prediction task may be constructed, which is convenient for downstream use. For example, the generated text is fed back to the user, or the user question is answered based on the generated text, which is not limited in the embodiment.

[0044] In practice, the multimedia processing model may adopt a hierarchical Transformer architecture to divide the byte sequence into segments to improve computing efficiency. The model may include three main components, namely a linear projection unit, a sequence decoding unit and a byte decoding unit, to complete prediction at a byte level.

[0045] Furthermore, after inputting the binary byte sequence into the multimedia processing model, in order to achieve prediction in a binary byte dimension, it is required to first perform byte to vector conversion. In this embodiment, the specific implementation is described as follows:

inputting the binary byte sequence into the linear projection unit, and converting, by the linear projection unit, binary bytes in the binary byte sequence into byte vectors; constructing an initial byte sequence vector corresponding to the binary byte sequence based on a byte vector corresponding to each of the binary bytes; and converting the initial byte sequence vector into a one-dimensional vector, and mapping the one-dimensional vector to a binary byte space to obtain the byte sequence vector.

[0046] Specifically, the byte vector refers to a vector expression obtained by converting a

binary byte in the binary bytes into a vector. Correspondingly, the initial byte sequence vector specifically refers to a vector expression obtained by concatenating the byte vector corresponding to each binary byte. The binary byte space specifically refers to mapping the one-dimensional vector to the byte space, so that the intrinsic correlation of the multimedia data can be reflected by the byte sequence vector mapped from the byte space, thus facilitating model prediction.

[0047] Based on this, after inputting the binary byte sequence into the multimedia processing model, the sequence may be first vector mapped by the linear projection unit. In this case, the binary bytes in the binary byte sequence may be converted into byte vectors by the linear projection unit. On this basis, the initial byte sequence vector corresponding to the binary byte sequence may be constructed based on the byte vector corresponding to each binary byte. Then, the initial byte sequence vector may be converted into a one-dimensional vector, and the one-dimensional vector may be mapped to the binary byte space to obtain the corresponding byte sequence vector of the multimedia resource, which may be used for subsequent prediction.

[0048] In practice, when mapping the binary byte sequence, the linear projection unit may use each binary byte sequence as a matrix, and each byte may be represented as a 257 dimensional vector through one hot encoding, including all 256 byte values and a<eop> (end of a sequence) marker. Afterwards, the linear projection unit converts the sequence into one-dimensional vectors and map the one-dimensional vectors into the binary byte space to obtain the byte sequence vectors.

[0049] In summary, by introducing the linear projection unit into the model to map the binary byte sequence to the binary byte space, so that the byte sequence vectors can be obtained while the data relationship in the multimedia data can still be reflected, thereby achieving accurate prediction of the model.

[0050] Furthermore, after obtaining the byte sequence vector, the byte prediction may be completed by the sequence decoding unit and byte decoding unit in the model. In this embodiment, the implementation is specifically described as follows:

inputting the byte sequence vector into the sequence decoding unit, and performing autoregressive prediction on the byte sequence vector by the sequence decoding unit to obtain the sequence feature; and inputting the sequence feature into the byte decoding unit, and

performing autoregressive reconstruction on the sequence feature by the byte decoding unit to obtain the prediction bytes.

[0051] Based on this, after obtaining the byte sequence vector by mapping of the linear projection unit, the byte sequence vector may be inputted to the sequence decoding unit, which performs autoregressive prediction on the byte sequence vector to achieve structural prediction at the sequence level to obtain the sequence feature. Afterwards, the sequence feature is inputted into the byte decoding unit, which performs autoregressive reconstruction on the sequence feature to achieve binary byte prediction at the byte dimension, to obtain the prediction bytes for subsequent conversion into the target task data for downstream task processing.

[0052] In practice, the sequence decoding unit can achieve autoregressive prediction of the features of subsequent sequences and learn the structure of data, and the byte decoding unit can reconstruct the byte sequence within the segment based on the segment features, thereby obtaining prediction characters for subsequent use.

[0053] In summary, by predicting features from the sequence dimension with a sequence decoder and generating the prediction bytes from the byte dimension based on the prediction result, it can be ensured that the model outputs more accurate prediction result from the binary dimension.

[0054] After obtaining the prediction bytes mentioned above, considering that the final result needs to be fed back to the user or business end for use, it is required to further convert the prediction bytes into target task data. In this embodiment, the implementation is specifically described as follows:

determining an inverse conversion strategy corresponding to the conversion strategy, and parsing the prediction bytes based on the inverse conversion strategy; and determining prediction code point information corresponding to the prediction bytes based on a parsing result, and generating the target task data corresponding to the target prediction task based on the prediction code point information.

[0055] Specifically, the inverse conversion strategy refers to a strategy of converting the prediction bytes into the target task data, that is, the prediction bytes may be converted into text, images, or audio, etc. through the inverse conversion strategy. Correspondingly, the

prediction code point information specifically refers to code point information corresponding to the prediction byte.

[0056] Based on this, after the model outputs the prediction bytes, the inverse conversion strategy corresponding to the conversion strategy may be determined first. Based on this, the prediction bytes may be parsed according to the inverse conversion strategy. In this way, the prediction code point information corresponding to the prediction byte may be determined based on a parsing result, and the target task data corresponding to the target prediction task may be generated based on the prediction code point information for downstream use.

[0057] Following the previous example, after obtaining the binary byte sequence corresponding to the question text "Who is the author of the ancient poem A", the binary byte sequence may be inputted into the multimedia processing model. Firstly, mapping is performed on the binary byte sequence by the linear projection unit in the multimedia processing model to obtain the byte sequence vector ZLc. Secondly, the byte sequence vector ZLc may be inputted into the sequence decoder in the multimedia processing model, and the sequence feature may be predicted through autoregression by the sequence decoder to obtain the sequence vector Lc. Finally, the sequence vector Lc may be inputted into the byte decoder in the multimedia processing model, and the byte sequence within the sequence may be reconstructed through autoregression by the byte decoder to generate the prediction bytes. After obtaining the prediction bytes, the prediction bytes are converted according to the reverse encoding process described above to obtain the answer text "A" corresponding to the question text "Who is the author of the ancient poem A", and the answer text is fed back to the user.

[0058] In summary, by using a multimedia processing model that includes the linear projection unit, the sequence decoding unit and the byte decoding unit to process the multimedia data, the model can provide a prediction result in the byte dimension, effectively improving the prediction accuracy and processing efficiency of the model. In addition, less computing resource will be occupied.

[0059] In addition, in order to enable the multimedia processing model to perform the prediction at the byte dimension, it is required to provide the model with the capability during the training phase. In this embodiment, the model is trained by:

acquiring sample data and a sample label corresponding to the sample data, converting

the sample data into a sample binary byte sequence, and converting the sample label into label bytes; inputting the sample binary byte sequence into an initial multimedia processing model for processing, to obtain training prediction bytes; and training the initial multimedia processing model based on the training prediction bytes and the label bytes until obtaining the multimedia processing model that meets the training stopping condition.

[0060] Specifically, the sample data and sample label refer to samples used to train the multimedia processing model. By converting the samples into the binary sequence, the model can learn the prediction capability of binary bytes, thereby improving the explore capability of the model on byte level data. Correspondingly, the label byte is the binary byte corresponding to the sample label. The training prediction byte specifically refers to a prediction byte outputted by the model. The training stopping condition specifically refers to a condition for stopping the training of the multimedia processing model, including but not limited to a loss value comparison condition, a validation condition, or an iteration number condition, which is not limited in the embodiment.

[0061] Based on this, during the model training phase, the sample data and the sample label corresponding to the sample data may be acquired first. Afterwards, the sample data may be converted into a sample binary byte sequence, and the sample label may be converted into label bytes. On this basis, the model may be trained by inputting the sample binary byte sequence into the initial multimedia processing model for processing, so that the training prediction bytes can be obtained. On this basis, the initial multimedia processing model is trained based on the training prediction bytes and label bytes until a multimedia processing model that meets the training stopping condition is obtained.

[0062] With the data processing method according to the embodiment, in order to improve the comprehensive understanding and simulation of digital systems by the model, so as to enhance the processing capability and efficiency of multimedia data, after obtaining the multimedia data of the target prediction task, the multimedia data may be first converted into a binary byte sequence. In this case, the binary byte sequence may be inputted into the multimedia processing model, which includes a linear projection unit, a sequence decoding unit and a byte decoding unit. On this basis, the binary byte sequence may be converted into a byte sequence vector by the linear projection unit, and the byte sequence vector is processed by the sequence decoding unit to obtain a sequence feature, thereby exploring a structural

relationship at a byte sequence dimension. The sequence feature is processed by the byte decoding unit, to obtain prediction bytes, thereby exploring a byte relationship at a byte dimension. Finally, target task data corresponding to the target prediction task may be constructed based on the prediction bytes, achieving processing at a binary byte dimension by the model. Therefore, the integration process between different types of data is simplified, improving the applicability and performance of the model in cross modal tasks.

[0063] In the following, the data processing method is further described in connection with Figure 3 by taking an application of the data processing method according to the present disclosure in an image data processing scenario as an example. Figure 3 shows a flowchart of a processing procedure of a data processing method according to an embodiment of the present disclosure, which specifically includes the following steps S302 to S320.

[0064] In step S302, multimedia data of a target prediction task is acquired, and code point information corresponding to the multimedia data is determined according to a conversion strategy associated with the target prediction task.

[0065] In step S304, the multimedia data is processed based on the code point information and an encoding rule configured in conversion strategy, to obtain a binary byte sequence corresponding to the multimedia data.

[0066] In step S306, the binary byte sequence is inputted into a multimedia processing model, which includes a linear projection unit, a sequence decoding unit and a byte decoding unit.

[0067] In step S308, binary bytes in the binary byte sequence are converted into byte vectors by the linear projection unit.

[0068] In step S310, an initial byte sequence vector corresponding to the binary byte sequence is constructed based on the byte vector corresponding to each binary byte.

[0069] In step S312, the initial byte sequence vector is converted into a one-dimensional vector, and the one-dimensional vector is mapped to the binary byte space to obtain a byte sequence vector.

[0070] In step S314, autoregressive prediction is performed on the byte sequence vector by the sequence decoding unit to obtain a sequence feature.

[0071] In step S316, the sequence feature is inputted into the byte decoding unit, and

autoregressive reconstruction is performed on the sequence feature by the byte decoding unit to obtain prediction bytes.

[0072] In step S318, an inverse conversion strategy corresponding to the conversion strategy is determined, and the prediction bytes are parsed based on the inverse conversion strategy.

[0073] In step S320, prediction code point information corresponding to the prediction byte is determined based on a parsing result, and the target task data corresponding to the target prediction task is generated based on the prediction code point information.

[0074] With the data processing method according to the embodiment, in order to improve the comprehensive understanding and simulation of digital systems by the model, so as to enhance the processing capability and efficiency of multimedia data, after obtaining the multimedia data of the target prediction task, the multimedia data may be first converted into a binary byte sequence. In this case, the binary byte sequence may be inputted into the multimedia processing model, which includes a linear projection unit, a sequence decoding unit and a byte decoding unit. On this basis, the binary byte sequence may be converted into a byte sequence vector by the linear projection unit, and the byte sequence vector is processed by the sequence decoding unit to obtain a sequence feature, thereby exploring a structural relationship at a byte sequence dimension. The sequence feature is processed by the byte decoding unit, to obtain prediction bytes, thereby exploring a byte relationship at a byte dimension. Finally, target task data corresponding to the target prediction task may be constructed based on the prediction bytes, achieving processing at a binary byte dimension by the model. Therefore, the integration process between different types of data is simplified, improving the applicability and performance of the model in cross modal tasks.

[0075] Corresponding to the above method embodiments, an embodiment of a data processing device is further provided according to the present disclosure. Figure 4 shows a schematic diagram of a structure of a data processing device according to embodiment of the present disclosure. As shown in Figure 4, the device includes an acquisition module 402, an input module 404, a conversion module 406 and a construction module 408. The acquisition module 402 is configured to acquire multimedia data of a target prediction task, and convert the multimedia data into a binary byte sequence. The input module 404 is configured to input the binary byte sequence into a multimedia processing model, where the multimedia

processing model includes a linear projection unit, a sequence decoding unit and a byte decoding unit. The conversion module 406 is configured to convert, by the linear projection unit, the binary byte sequence into a byte sequence vector, and process the byte sequence vector by the sequence decoding unit to obtain a sequence feature. The construction module 408 is configured to process the sequence feature by the byte decoding unit to obtain prediction bytes, and construct target task data corresponding to the target prediction task based on the prediction bytes.

[0076] In a preferred embodiment, the acquisition module 402 is further configured to:

determine code point information corresponding to the multimedia data based on a conversion strategy associated with the target prediction task; and process the multimedia data based on the code point information and an encoding rule configured in the conversion strategy, to obtain the binary byte sequence corresponding to the multimedia data.

[0077] In a preferred embodiment, the conversion module 406 is further configured to:

input the binary byte sequence into the linear projection unit, and convert, by the linear projection unit, binary bytes in the binary byte sequence into byte vectors; construct an initial byte sequence vector corresponding to the binary byte sequence based on a byte vector corresponding to each of the binary bytes; and convert the initial byte sequence vector into a one-dimensional vector, and map the one-dimensional vector to a binary byte space to obtain the byte sequence vector.

[0078] In a preferred embodiment, the conversion module 406 is further configured to:

input the byte sequence vector into the sequence decoding unit, and perform autoregressive prediction on the byte sequence vector by the sequence decoding unit to obtain the sequence feature,

where the construction module 408 is further configured to:

input the sequence feature into the byte decoding unit, and perform autoregressive reconstruction on the sequence feature by the byte decoding unit to obtain the prediction bytes.

[0079] In a preferred embodiment, the construction module 408 is further configured to:

determine an inverse conversion strategy corresponding to the conversion strategy, and

parsing the prediction bytes based on the inverse conversion strategy; and determine prediction code point information corresponding to the prediction bytes based on a parsing result, and generate the target task data corresponding to the target prediction task based on the prediction code point information.

5 **[0080]** In a preferred embodiment, the device further includes a training module. The training module is configured to: acquire sample data and a sample label corresponding to the sample data, convert the sample data into a sample binary byte sequence, and convert the sample label into label bytes; input the sample binary byte sequence into an initial multimedia processing model for processing, to obtain training prediction bytes; and train the initial
10 multimedia processing model based on the training prediction bytes and the label bytes until the multimedia processing model that meets the training stopping condition is obtained.

[0081] With the data processing device according to the embodiment, in order to improve the comprehensive understanding and simulation of digital systems by the model, so as to enhance the processing capability and efficiency of multimedia data, after obtaining the
15 multimedia data of the target prediction task, the multimedia data may be first converted into a binary byte sequence. In this case, the binary byte sequence may be inputted into the multimedia processing model, which includes a linear projection unit, a sequence decoding unit and a byte decoding unit. On this basis, the binary byte sequence may be converted into a byte sequence vector by the linear projection unit, and the byte sequence vector is processed
20 by the sequence decoding unit to obtain a sequence feature, thereby exploring a structural relationship at a byte sequence dimension. The sequence feature is processed by the byte decoding unit, to obtain prediction bytes, thereby exploring a byte relationship at a byte dimension. Finally, target task data corresponding to the target prediction task may be constructed based on the prediction bytes, achieving processing at a binary byte dimension by
25 the model. Therefore, the integration process between different types of data is simplified, improving the applicability and performance of the model in cross modal tasks.

[0082] The above is a schematic solution of the data processing device for 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
30 of the technical solution of the data processing device that are not described in detail, please refer to the description of the technical solution of the data processing method mentioned

above.

[0083] Figure 5 illustrates a block diagram of a structure of a computing device 500 according to an embodiment of the present disclosure. The components of the computing device 500 include, but are not limited to, a memory 510 and a processor 520. The processor 520 is connected to the memory 510 via a bus 530, and a database 550 is used to store data.

[0084] The computing device 500 further includes an access device 540, the access device 540 enables the computing device 500 to communicate via one or more networks 560. 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 540 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).

[0085] In an embodiment of the present disclosure, the above components of the computing device 500, as well as other components not shown in Figure 5, 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 5 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.

[0086] The computing device 500 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 500 may also be a mobile or stationary server.

[0087] The processor 520 is used to execute computer executable instructions, which, when executed by the processor, perform the steps of the data processing method described above.

[0088] 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 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 described above.

[0089] A computer-readable storage medium is further 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 described above.

[0090] 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 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 described above.

[0091] A computer program product is further 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 described above.

[0092] 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 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 technical solution of the data processing method described above.

[0093] 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 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.

[0094] 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 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.

[0095] 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.

[0096] 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 may be referred to the relevant description of other embodiments.

[0097] 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:

acquiring multimedia data of a target prediction task, and converting the multimedia data
5 into a binary byte sequence;

inputting the binary byte sequence into a multimedia processing model, wherein the multimedia processing model comprises a linear projection unit, a sequence decoding unit and a byte decoding unit;

converting, by the linear projection unit, the binary byte sequence into a byte sequence
10 vector, and processing the byte sequence vector by the sequence decoding unit to obtain a sequence feature; and

processing the sequence feature by the byte decoding unit to obtain prediction bytes, and constructing target task data corresponding to the target prediction task based on the prediction bytes.

15 2. The data processing method according to claim 1, wherein the converting the multimedia data into a binary byte sequence comprises:

determining code point information corresponding to the multimedia data based on a conversion strategy associated with the target prediction task; and

processing the multimedia data based on the code point information and an encoding rule
20 configured in the conversion strategy, to obtain the binary byte sequence corresponding to the multimedia data.

3. The data processing method according to claim 1, wherein the converting, by the linear projection unit, the binary byte sequence into a byte sequence vector comprises:

inputting the binary byte sequence into the linear projection unit, and converting, by the
25 linear projection unit, binary bytes in the binary byte sequence into byte vectors;

constructing an initial byte sequence vector corresponding to the binary byte sequence based on a byte vector corresponding to each of the binary bytes; and

converting the initial byte sequence vector into a one-dimensional vector, and mapping the one-dimensional vector to a binary byte space to obtain the byte sequence vector.

4. The data processing method according to claim 1, wherein the processing the byte sequence vector by the sequence decoding unit to obtain a sequence feature comprises:

inputting the byte sequence vector into the sequence decoding unit, and performing autoregressive prediction on the byte sequence vector by the sequence decoding unit to obtain the sequence feature,

wherein the processing the sequence feature by the byte decoding unit to obtain prediction bytes comprises:

inputting the sequence feature into the byte decoding unit, and performing autoregressive reconstruction on the sequence feature by the byte decoding unit to obtain the prediction bytes.

5. The data processing method according to claim 2, wherein the constructing target task data corresponding to the target prediction task based on the prediction bytes comprises:

determining an inverse conversion strategy corresponding to the conversion strategy, and parsing the prediction bytes based on the inverse conversion strategy; and

determining prediction code point information corresponding to the prediction bytes based on a parsing result, and generating the target task data corresponding to the target prediction task based on the prediction code point information.

6. The data processing method according to any one of claims 1 to 5, wherein before the inputting the binary byte sequence into a multimedia processing model, the method further comprises:

acquiring sample data and a sample label corresponding to the sample data, converting the sample data into a sample binary byte sequence, and converting the sample label into label bytes;

inputting the sample binary byte sequence into an initial multimedia processing model for processing, to obtain training prediction bytes; and

training the initial multimedia processing model based on the training prediction bytes and the label bytes until obtaining the multimedia processing model that meets the training stopping condition.

7. A data processing device, comprising:

an acquisition module, configured to acquire multimedia data of a target prediction task, and convert the multimedia data into a binary byte sequence;

an input module, configured to input the binary byte sequence into a multimedia processing model, wherein the multimedia processing model comprises a linear projection unit, a sequence decoding unit and a byte decoding unit;

a conversion module, configured to convert, by the linear projection unit, the binary byte sequence into a byte sequence vector, and process the byte sequence vector by the sequence decoding unit to obtain a sequence feature; and

a construction module, configured to process the sequence feature by the byte decoding unit to obtain prediction bytes, and construct target task data corresponding to the target prediction task based on the prediction bytes.

8. A computing device, comprising:

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 6.

9. A computer-readable storage medium, comprising computer executable instructions 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 6.

10. 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 6.