

METHODS, APPARATUSES, AND COMPUTING DEVICES FOR TRAININGS OF LEARNING MODELS

TECHNICAL FIELD

[0001] Embodiments of the present specification relate to the field of machine
5 learning, and specifically, to methods, apparatuses, and computing devices for
trainings of learning models.

BACKGROUND

[0002] Deep learning originates from the research of artificial neural networks, it's
a new field of machine learning research in recent years, and has wide application
10 prospects in various fields. Currently, in most deep learning methods, a deep learning
model (which can be also referred to as a deep neural network) is usually trained in a
batch learning way. In this way, training data sets usually need to be prepared before a
learning task starts (i.e., offline training).

[0003] However, in many current applications, data may arrive successively in a
15 form of stream, and therefore the batch learning way may not be very practical for
such applications.

SUMMARY

[0004] Considering the previously described problem in the existing technology,
embodiments of the present specification provide methods, apparatuses, and
20 computing devices for trainings of learning models.

[0005] According to an aspect, embodiments of the present specification provide a
method for trainings of learning models, including the following: current streaming
sample data is received; and a current deep learning model is trained based on the
current streaming sample data, where parameters of a shallow learning model are used
25 as initialization parameters of the current deep learning model, and the shallow
learning model is obtained through training based on historical sample data associated
with the current streaming sample data.

[0006] According to another aspect, embodiments of the present specification provide an apparatus for trainings of learning models, including: a receiving unit, configured to receive current streaming sample data; and a training unit, configured to train a current deep learning model based on the current streaming sample data, where parameters of a shallow learning model are used as initialization parameters of the current deep learning model, and the shallow learning model is obtained through training based on historical sample data associated with the current streaming sample data.

[0007] According to another aspect, embodiments of the present specification provide a computing device, including at least one processor; and a storage that communicates with the at least one processor, where an executable instruction is stored in the memory, and the at least one processor is enabled to implement the previously described method when the executable instruction is executed by the at least one processor.

[0008] It can be seen that in the technical solutions, when the current deep learning model is trained based on the current streaming sample data, the parameters of the trained shallow learning model are used as the initialization parameters of the current deep learning model, so that convergence of the deep learning model can be accelerated to efficiently finish a model training process, and performance of the deep learning model can also be improved.

BRIEF DESCRIPTION OF DRAWINGS

[0009] The following further describes the previously described characteristics, technical features, advantages, and implementations of the present invention in a clear and easy-to-understand way by describing preferred embodiments and with reference to the accompanying drawings:

[0010] FIG. 1 is a schematic flowchart illustrating a method for trainings of learning models according to an embodiment;

[0011] FIG. 2A illustrates an example training process of a shallow learning model according to an embodiment;

[0012] FIG. 2B illustrates an example training process of a deep learning model according to an embodiment;

[0013] FIG. 3 is a schematic block diagram illustrating an apparatus for trainings of learning models according to an embodiment; and

[0014] FIG. 4 is a hardware structural diagram illustrating a computing device for trainings of learning models according to an embodiment.

5

DESCRIPTION OF IMPLEMENTATIONS

[0015] At present, the subject described here is discussed with reference to embodiments. It is worthwhile to understand that the discussion of these embodiments is merely intended to enable a person skilled in the art to better understand and implement the subject described here, and is not intended to limit the protection scope, applicability, or examples described in the claims. The functions and arrangements of the discussed elements can be changed without departing from the protection scope of the claims. In the embodiments, various processes or components can be omitted, replaced, or added as required.

[0016] Currently, a deep learning model is usually trained in a batch learning way. In the batch learning way, training data sets usually need to be prepared before a learning task starts, and then these training data sets are used to train the deep learning model (i.e., offline training). However, in some applications, data may arrive successively in a form of stream, and therefore the batch learning way may not be very suitable for such applications. In addition, in the batch learning way, the training data sets prepared in advance may occupy relatively large storage space, and therefore the batch learning way is also not suitable for some scenarios with limited storage space.

[0017] The embodiments of the present specification provide technical solutions for training a learning model based on streaming sample data. In the present specification, the streaming sample data usually can include sample data continuously generated from a data sample source, for example, log files, online shopping data, game player activity data, and social networking site information data generated by a web application. The streaming sample data can also be referred to as real-time sample data whose time span is usually between hundreds of milliseconds and several seconds. Model training performed based on the streaming sample data usually can also be considered as online learning.

[0018] Specifically, in the technical solutions of the present specification, current streaming sample data can be received. Then, a current deep learning model can be trained based on the current streaming sample data. Parameters of a shallow learning model can be used as initialization parameters of the current deep learning model, and
5 the shallow learning model can be obtained through training based on historical sample data associated with the current streaming sample data.

[0019] It can be seen that in the technical solutions, when the current deep learning model is trained based on the current streaming sample data, the parameters of the trained shallow learning model are used as the initialization parameters of the current
10 deep learning model, so that convergence of the deep learning model can be accelerated to efficiently finish a model training process, and performance of the deep learning model can also be improved.

[0020] In addition, in the technical solutions, the deep learning model is trained based on the streaming sample data, so that no training data set needs to be prepared
15 in advance, thereby effectively saving storage space.

[0021] In some cases, data may show a phenomenon of concept shift, that is, the statistical characteristic of the data may change in an unforeseeable way over time. In these cases, the deep learning model is trained based on the streaming sample data, so that the deep learning model can be adjusted in time as the data changes, thereby
20 improving a prediction effect of the deep learning model and achieving good scalability.

[0022] In the present specification, a structure of the shallow learning model can be simpler than a structure of the deep learning model. For example, the shallow learning model can have no hidden layer, or have fewer hidden layers than the deep
25 learning model.

[0023] For example, the shallow learning model can be a logistic regression (LR) model, and the deep learning model can have one or more hidden layers. In some cases, for ease of implementation, the deep learning model can be initially constructed as having one hidden layer.

30 [0024] The following describes the technical solutions in detail with reference to specific embodiments.

[0025] FIG. 1 is a schematic flowchart illustrating a method for trainings of learning models according to an embodiment.

[0026] As shown in FIG. 1, in step 102, current streaming sample data can be received.

[0027] In step 104, a current deep learning model can be trained based on the current streaming sample data.

5 **[0028]** Parameters of a shallow learning model can be used as initialization parameters of the current deep learning model, and the shallow learning model can be obtained through training based on historical sample data associated with the current streaming sample data.

10 **[0029]** For example, the association between the historical sample data and the current streaming sample data can indicate that the historical sample data has one or more sample features that are the same as the current streaming sample data.

[0030] In an embodiment, the historical sample data can be historical streaming sample data before the current streaming sample data. For example, the historical streaming sample data may include one or more batches of streaming sample data
15 before the current streaming sample data. In this case, the shallow learning model can be obtained through online training based on the historical streaming sample data.

[0031] In an embodiment, the historical sample data can be offline sample data associated with the current streaming sample data. For example, the offline sample data can be data having one or more sample features that are the same as the current
20 streaming sample data. In this case, the shallow learning model can be obtained through offline training based on the offline sample data.

[0032] It can be seen that in the present specification, the shallow learning model can be obtained through training online or offline, to satisfy different application needs.

25 **[0033]** In an embodiment, when the parameters of the shallow learning model are used as the initialization parameters of the current deep learning model, parameters of layers of the shallow learning model can be used as initialization parameters of corresponding layers of the current deep learning model. For example, the parameters of the layers of the shallow learning model can be mapped to the corresponding layers
30 of the current deep learning model in a one-to-one mapping way. It helps accelerate convergence of the deep learning model to shorten training time, and also helps improve performance of the deep learning model. In addition, parameters of the remaining layers of the current deep learning model can be initialized in a randomized way.

[0034] In an embodiment, in step 104, the current deep learning model can be trained by using various methods applicable to the art, such as online gradient descent.

[0035] In an embodiment, after the training of the current deep learning model is finished to obtain a trained deep learning model, performance of the trained deep learning model can be compared with performance of the current deep learning model.

[0036] For example, indicators used to evaluate performance can include various applicable indicators such as an area under curve (AUC), accuracy, coverage, and an F1 score, and are not limited in the present specification.

[0037] In response to determining that the performance of the trained deep learning model is improved compared to the performance of the current deep learning model, for example, AUC of the trained deep learning model is higher than AUC of the current deep learning model, the trained deep learning model can be used as the latest deep learning model; or

[0038] in response to determining that the performance of the trained deep learning model is not improved compared to the performance of the current deep learning model, for example, AUC of the trained deep learning model is lower than or basically approximates to AUC of the current deep learning model, the current deep learning model can be considered to be updated.

[0039] For example, the quantity of hidden layers of the current deep learning model can be increased to obtain a deep learning model with the increased quantity of layers. For example, the quantity of hidden layers of the current deep learning model can be increased by 1 or more, which can be determined based on actual needs, etc.

[0040] Then, the deep learning model with the increased quantity of layers can be trained based on the current streaming sample data, to obtain a new deep learning model.

[0041] In an embodiment, the parameters of the shallow learning model can be used as initialization parameters of the deep learning model with the increased quantity of layers. It can help accelerate convergence of the deep learning model with the increased quantity of layers to efficiently obtain the new deep learning model, and also help improve performance of the new deep learning model.

[0042] Then, the latest deep learning model can be determined based on a result of performance comparison between the new deep learning model and the current deep learning model.

[0043] For example, in response to determining that the performance of the new deep learning model is improved compared to the performance of the current deep learning model, the new deep learning model can be used as the latest deep learning model; or in response to determining that the performance of the new deep learning model is not improved compared the performance of the current deep learning model, the current deep learning model can be used as the latest deep learning model.

[0044] It can be seen that in the previously described way, a current optimal deep learning model can be effectively selected as the latest deep learning model.

[0045] In an embodiment, the latest deep learning model can be directly used as the latest learning model to be applied.

[0046] Alternatively, in another embodiment, the latest deep learning model and the shallow learning model can be weighted to obtain the latest learning model to be applied.

[0047] For example, the latest deep learning model and the shallow learning model can respectively have corresponding weights. Their weights can be predefined, or can be defined by a user based on actual needs. For example, a weight of the latest deep learning model can be 70% and a weight of the shallow learning model can be 30%; or a weight of the latest deep learning model can be 100% and a weight of the shallow learning model can be 0. As such, the method is applicable to different application scenarios.

[0048] The following describes the technical solutions in the present specification with reference to specific examples. It is worthwhile to understand that these examples are merely intended to help a person skilled in the art better understand the embodiments, but are not intended to limit the scope thereof.

[0049] FIG. 2A illustrates an example training process of a shallow learning model according to an embodiment.

[0050] As shown in FIG. 2A, in step 202A, historical sample data can be obtained.

[0051] In step 204A, shallow learning model training can be performed based on the historical sample data to obtain a shallow learning model.

[0052] For example, the historical sample data can be one or more batches of streaming sample data before the previously described current streaming sample. In this case, shallow learning model training can be performed each time one batch of streaming sample data is received. As such, one or more batches of streaming sample

data may be required to complete shallow learning model training online, to obtain the shallow learning model.

5 **[0053]** For another example, the historical sample data can be offline sample data associated with the previously described current streaming sample. In this case, shallow learning model training can be performed offline based on the offline sample data, to obtain the shallow learning model.

10 **[0054]** It can be understood that the process shown in FIG. 2A can occur before a training process of a deep learning model (i.e., the process shown in FIG. 2B), so that parameters of the shallow learning model can be used as initialization parameters of a deep learning model.

[0055] FIG. 2B illustrates an example training process of a deep learning model according to an embodiment.

[0056] As shown in FIG. 2B, in step 202B, current streaming sample data can be received.

15 **[0057]** In step 204B, a current deep learning model can be trained based on the current streaming sample data, to obtain a trained deep learning model.

[0058] For example, during training, the current deep learning model can be first initialized. For example, parameters of layers of a shallow learning model can be obtained, and then the parameters of the layers of the shallow learning model can be used as initialization parameters of corresponding layers of the current deep learning model. Parameters of the remaining layers of the current deep learning model can be initialized in a randomized way.

20 **[0059]** In step 206B, it can be determined whether performance of the trained deep learning model is improved compared to performance of the current deep learning model.

[0060] In response to determining that the performance of the trained deep learning model is improved compared to the performance of the current deep learning model, in step 208B, the trained deep learning model can be used as the latest deep learning model; or

30 **[0061]** in response to determining that the performance of the trained deep learning model is not improved compared to the performance of the current deep learning model, in step 210B, the quantity of hidden layers of the current deep learning model can be increased, for example, can be increased by 1 or more.

[0062] In step 212B, a deep learning model with the increased quantity of layers can be trained to obtain a new deep learning model.

[0063] For example, during training of the deep learning model with the increased quantity of layers, the parameters of the layers of the shallow learning model can be used as initialization parameters of corresponding layers of the deep learning model with the increased quantity of layers, and parameters of the remaining layers of the deep learning model with the increased quantity of layers can be initialized in a randomized way.

[0064] In step 214B, it can be determined whether performance of the new deep learning model is improved compared to the performance of the current deep learning model.

[0065] In response to determining that the performance of the new deep learning model is improved compared to the performance of the current deep learning model, in step 216B, the new deep learning model can be used as the latest deep learning model; or

[0066] in response to determining that the performance of the new deep learning model is not improved compared to the performance of the current deep learning model, in step 218B, the current deep learning model can be used as the latest deep learning model.

[0067] It is worthwhile to understand that sequences of the steps shown in FIG. 2A and FIG. 2B are merely examples. The sequences of these steps can be changed correspondingly based on actual application or different design logic.

[0068] FIG. 3 is a schematic block diagram illustrating an apparatus for trainings of learning models according to an embodiment.

[0069] As shown in FIG. 3, an apparatus 300 can include a receiving unit 302 and a training unit 304.

[0070] The receiving unit 302 can be configured to receive current streaming sample data. The training unit 304 can be configured to train a current deep learning model based on the current streaming sample data. Parameters of a shallow learning model can be used as initialization parameters of the current deep learning model, and the shallow learning model can be obtained through training based on historical sample data associated with the current streaming sample data.

[0071] It can be seen that in the technical solutions, the parameters of the trained shallow learning model are used as the initialization parameters of the current deep

learning model, so that when the current deep learning model is trained based on the current streaming sample data, convergence of the deep learning model can be accelerated to efficiently finish a model training process, and performance of the deep learning model can also be improved.

5 **[0072]** In an embodiment, the historical sample data can be historical streaming sample data before the current streaming sample data. In this case, the shallow learning model can be obtained through online training based on the historical streaming sample data.

10 **[0073]** In an embodiment, the historical sample data can be offline sample data. In this case, the shallow learning model can be obtained through offline training based on the offline sample data.

[0074] In an embodiment, the apparatus 300 can further include an evaluation unit 306.

15 **[0075]** After the training unit 304 finishes the training of the current deep learning model to obtain a trained deep learning model, the evaluation unit 306 can be configured to evaluate performance of the trained deep learning model.

20 **[0076]** In response to determining that the performance of the trained deep learning model is improved compared to performance of the current deep learning model, the evaluation unit 306 can be configured to use the trained deep learning model as the latest deep learning model; or

25 **[0077]** in response to determining that the performance of the trained deep learning model is not improved compared to performance of the current deep learning model, the training unit 304 can be configured to increase the quantity of hidden layers of the current deep learning model to obtain a deep learning model with the increased quantity of layers, and the training unit 304 can be configured to train the deep learning model with the increased quantity of layers based on the current streaming sample data, to obtain a new deep learning model; and then the evaluation unit 306 can be configured to determine the latest deep learning model based on a result of performance comparison between the new deep learning model and the current deep learning model.

30 **[0078]** In an embodiment, the parameters of the shallow learning model can be used as initialization parameters of the deep learning model with the increased quantity of layers.

[0079] In an embodiment, in response to determining that performance of the new deep learning model is improved compared to the performance of the current deep learning model, the evaluation unit 306 can be configured to use the new deep learning model as the latest deep learning model; or

5 **[0080]** in response to determining that performance of the new deep learning model is not improved compared to the performance of the current deep learning model, the evaluation unit 306 can be configured to use the current deep learning model as the latest deep learning model.

[0081] In an embodiment, the apparatus 300 can further include a weighting unit 10 308. The weighting unit 308 can be configured to weight the latest deep learning model and the shallow learning model to obtain the latest learning model.

[0082] The units of the apparatus 300 can perform corresponding steps in the method embodiments shown in FIG. 1 to FIG. 2B. Therefore, for brevity of description, specific operations and functions of the units of the apparatus 300 are not 15 described here.

[0083] The apparatus 300 can be implemented by using hardware, software, or a combination of software and hardware. For example, when the apparatus 300 is implemented by using software, the apparatus 300 can be implemented by reading a corresponding executable instruction from a storage (for example, a non-volatile 20 storage) to a memory by using a processor of a device that the apparatus 300 is located in.

[0084] FIG. 4 is a hardware structural diagram illustrating a computing device for trainings of learning models according to an embodiment. As shown in FIG. 4, a computing device 400 can include at least one processor 402, a storage 404, a 25 memory 406, and a communications interface 408, and the at least one processor 402, the storage 404, the memory 406, and the communications interface 408 are connected together via a bus 410. The at least one processor 402 executes at least one executable instruction (i.e., the previously described element implemented in a form of software) stored or coded in the storage 404.

30 **[0085]** In an embodiment, when the executable instruction stored in the storage 404 is executed by the at least one processor 402, the computing device is enabled to implement the processes previously described with reference to FIG. 1 to FIG. 2B.

[0086] The computing device 400 can be implemented in any form applicable to the art, for example, includes but is not limited to a desktop computer, a laptop

computer, a smartphone, a tablet computer, a consumer electronic device, a wearable intelligent device, etc.

5 **[0087]** An embodiment of the present specification further provides a machine-readable storage medium. The machine-readable storage medium can store an executable instruction. When the executable instruction is executed by a machine, the machine is enabled to implement the specific processes of the method embodiments previously described with reference to FIG. 1 to FIG. 2B.

10 **[0088]** For example, the machine-readable storage medium can include but is not limited to a random access memory (Random Access Memory, RAM), a read-only memory (Read-Only Memory, ROM), an electrically-erasable programmable read-only memory (Electrically-Erasable Programmable Read-Only Memory, EEPROM), a static random access memory (Static Random Access Memory, SRAM), a hard disk, a flash memory, etc.

15 **[0089]** It is worthwhile to understand that the embodiments of the present specification are all described in a progressive way, for same or similar parts in the implementations, references can be made to each other, and each implementation focuses on a difference from the other embodiments. For example, the previously described apparatus embodiment, computing device embodiment, and machine-readable storage medium embodiment are basically similar to the method
20 embodiments, and therefore are described relatively simple. For related parts, references can be made to some descriptions of the method embodiments.

25 **[0090]** The specific embodiments of the present specification are described above. Other embodiments fall within the scope of the appended claims. In some cases, the actions or steps described in the claims can be performed based on sequences different from the sequences in the embodiments and can still achieve the desired results. In addition, the processes depicted in the accompanying drawings do not necessarily require the shown particular sequences or orders to achieve the desired results. In some implementations, multi-task processing and parallel processing can or may be advantageous.

30 **[0091]** It is worthwhile to understand that, for a person of ordinary skill in the art, various modifications to the embodiments in the present specification will be clear, and the general principle described here can be applied to other variants without departing from the protection scope of the claims.

CLAIMS

What is claimed is:

1. A method for trainings of learning models, comprising:
receiving current streaming sample data; and
5 training a current deep learning model based on the current streaming sample data, wherein parameters of a shallow learning model are used as initialization parameters of the current deep learning model, and the shallow learning model is obtained through training based on historical sample data associated with the current streaming sample data.
- 10 2. The method according to claim 1, wherein the historical sample data is historical streaming sample data before the current streaming sample data, and the shallow learning model is obtained through online training based on the historical streaming sample data.
3. The method according to claim 1, wherein the historical sample data is offline
15 sample data, and the shallow learning model is obtained through offline training based on the offline sample data.
4. The method according to any one of claims 1 to 3, further comprising:
after the training of the current deep learning model is finished to obtain a trained
deep learning model:
20 in response to determining that performance of the trained deep learning model is improved compared to performance of the current deep learning model, using the trained deep learning model as the latest deep learning model; or
in response to determining that performance of the trained deep learning model is not improved compared to performance of the current deep learning model,
25 performing the following operations:
increasing the quantity of hidden layers of the current deep learning model to obtain a deep learning model with the increased quantity of layers;
training the deep learning model with the increased quantity of layers based on the current streaming sample data, to obtain a new deep learning model; and
30 determining the latest deep learning model based on a result of performance comparison between the new deep learning model and the current deep learning model.
5. The method according to claim 4, wherein the parameters of the shallow learning model are used as initialization parameters of the deep learning model with

the increased quantity of layers.

6. The method according to claim 4 or 5, wherein the determining the latest deep learning model based on a result of performance comparison between the new deep learning model and the trained deep learning model comprises:

5 in response to determining that performance of the new deep learning model is improved compared to the performance of the current deep learning model, using the new deep learning model as the latest deep learning model; or

 in response to determining that performance of the new deep learning model is not improved compared to the performance of the current deep learning model,
10 using the current deep learning model as the latest deep learning model.

7. The method according to any one of claims 4 to 6, further comprising:

 weighting the latest deep learning model and the shallow learning model to obtain the latest learning model.

8. An apparatus for trainings of learning models, comprising:

15 a receiving unit, configured to receive current streaming sample data; and

 a training unit, configured to train a current deep learning model based on the current streaming sample data, wherein parameters of a shallow learning model are used as initialization parameters of the current deep learning model, and the shallow learning model is obtained through training based on historical sample data associated
20 with the current streaming sample data.

9. The apparatus according to claim 8, wherein the historical sample data is historical streaming sample data before the current streaming sample data, and the shallow learning model is obtained through online training based on the historical streaming sample data.

25 10. The apparatus according to claim 8, wherein the historical sample data is offline sample data, and the shallow learning model is obtained through offline training based on the offline sample data.

11. The apparatus according to any one of claims 8 to 10, further comprising an evaluation unit; and

30 after the training unit finishes the training of the current deep learning model to obtain a trained deep learning model:

 in response to determining that performance of the trained deep learning model is improved compared to performance of the current deep learning model, the evaluation unit is configured to use the trained deep learning model as the latest deep learning

model; or

in response to determining that performance of the trained deep learning model is not improved compared to performance of the current deep learning model:

5 the training unit is further configured to increase the quantity of hidden layers of the current deep learning model to obtain a deep learning model with the increased quantity of layers; and train the deep learning model with the increased quantity of layers based on the current streaming sample data, to obtain a new deep learning model; and

10 the evaluation unit is configured to determine the latest deep learning model based on a result of performance comparison between the new deep learning model and the current deep learning model.

12. The apparatus according to claim 11, wherein the parameters of the shallow learning model are used as initialization parameters of the deep learning model with the increased quantity of layers.

15 13. The apparatus according to claim 11 or 12, wherein the evaluation unit is specifically configured to:

in response to determining that performance of the new deep learning model is improved compared to the performance of the current deep learning model, use the new deep learning model as the latest deep learning model; or

20 in response to determining that performance of the new deep learning model is not improved compared to the performance of the current deep learning model, use the current deep learning model as the latest deep learning model.

14. The apparatus according to any one of claims 11 to 13, further comprising:

25 a weighting unit, configured to weight the latest deep learning model and the shallow learning model to obtain the latest learning model.

15. A computing device, comprising:

at least one processor; and

30 a storage that communicates with the at least one processor, wherein an executable instruction is stored in the memory, and the at least one processor is enabled to implement the method according to any one of claims 1 to 7 when the executable instruction is executed by the at least one processor.