

RECOMMENDATION METHOD AND DEVICE EMPLOYING DEEP LEARNING, ELECTRONIC APPARATUS, MEDIUM, AND PROGRAM

5

CROSS-REFERENCE TO RELATED APPLICATION

The present application claims priority to Chinese Patent Application No.
10 CN201810024218.9 filed on January 10, 2018, and entitled "RECOMMENDATION
METHOD AND DEVICE EMPLOYING DEEP LEARNING, ELECTRONIC
APPARATUS, MEDIUM, AND PROGRAM", the disclosure of which is incorporated
herein by reference in its entirety.

15

TECHNICAL FIELD

The present disclosure relates to Internet technologies, and in particular, to a
method and apparatus for deep learning-based recommendation, an electronic device,
20 a medium, and a program.

BACKGROUND

25 Recommendation systems have become the core of e-commerce and many
online sharing services, and technically, the recommendation systems are generally
generated based on content or collaborative filtering. In the aspect of the content, the
recommendation systems are mainly based on characteristics of items, and in the
aspect of the collaborative filtering, the recommendation systems are based on users
30 having similar histories; in practical application, the collaborative filtering has already
shown great performance.

SUMMARY

35

Embodiments of the present disclosure provide deep learning-based
recommendation technologies.

According to one aspect of embodiments of the present disclosure, provided is a method for deep learning-based recommendation, including: respectively obtaining related information of a target user and related information of a target item; respectively using at least two sub-models in an integrated model to obtain, based on
5 the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models, the operating probability being used for indicating a probability, obtained based on one sub-model, about the target user operating the target item, and the integrated model including at least two sub-models; obtaining, based on the operating probabilities corresponding to
10 the at least two sub-models, a target probability about the target user operating the target item; and recommending the target item to the target user based on the target probability.

In another embodiment based on the method of the present application, the related information of the target user includes: history information about a historical
15 item being operated by the target user, the historical item including at least one item operated by the target user;

and/or

the related information of the target item includes history information about the target item being operated by a historical user, the historical user including at least
20 one user who operated the target item.

In another embodiment based on the method of the present application, the respectively using at least two sub-models in an integrated model to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models includes:

25 using the sub-models to respectively vectorize the related information of the target user and the related information of the target item to obtain user vectors and item vectors; and

obtaining the operating probabilities based on the user vectors and the item vectors.

30 In another embodiment based on the method of the present application, the obtaining the operating probabilities based on the user vectors and the item vectors includes:

calculating dot products of the user vectors and the item vectors to obtain

operate values, and obtaining the operating probabilities based on the operate values.

In another embodiment based on the method of the present application, the obtaining, based on the operating probabilities corresponding to the at least two sub-models, a target probability about the target user operating the target item includes:

5 respectively calculating products of the operating probabilities and model weight values of the sub-models corresponding to the operating probabilities, each model weight value corresponding to one sub-model; and

 adding the products corresponding to the at least two operating probabilities to obtain the target probability about the target user operating the target item.

10 In another embodiment based on the method of the present application, the recommending the target item to the target user based on the target probability includes:

 in response to the target probability being greater than or equal to a preset threshold, recommending the target item corresponding to the target probability to the
15 target user; and

 in response to the target probability being less than a preset threshold, recommending no target item corresponding to the target probability to the target user.

 In another embodiment based on the method of the present application, the integrated model is obtained by training a relation between the at least two sub-
20 models.

 In another embodiment based on the method of the present application, a training mode for the integrated model includes:

 training the integrated model based on a historical user and a historical item, where a historical probability about the historical user operating the historical item is
25 known.

 In another embodiment based on the method of the present application, the training the integrated model based on a historical user and a historical item includes:

 establishing a historical vector corresponding to the historical user and the historical item based on the historical user and the historical item;

30 decomposing the historical vector using the integrated model to obtain at least one sample user vector and at least one sample item vector;

 obtaining a predicted probability based on the sample user vector and the sample item vector, the predicted probability indicating the probability about the

historical user operating the historical item predicted by the integrated model; and
training the integrated model based on the predicted probability and the
historical probability.

In another embodiment based on the method of the present application, the
5 decomposing the historical vector using the integrated model to obtain at least one
sample user vector and at least one sample item vector includes:

decomposing the historical vector using an initial integrated model to obtain at
least one sample user vector and at least one sample item vector, the initial integrated
model including one sub-model.

10 In another embodiment based on the method of the present application, the
obtaining a predicted probability based on the sample user vector and the sample item
vector includes:

calculating a dot product of the sample user vector and the sample item vector
using the initial integrated model to obtain a predicted value about the historical user
15 clicking the historical item, and obtaining the predicted probability based on the
predicted value.

In another embodiment based on the method of the present application, the
training the integrated model based on the predicted probability and the historical
probability includes:

20 allocating, according to a difference between the predicted probability and the
historical probability, data weight values for the historical user and the historical item
corresponding to the predicted probability, where the greater the difference is, the
greater the data weight values are; and

iteratively training the integrated model based on the sample user vector and
25 the sample item vector combined with the data weight values.

In another embodiment based on the method of the present application, the
iteratively training the integrated model based on the sample user vector and the
sample item vector combined with the data weight values includes:

adding one sub-model to the initial integrated model to constitute a first
30 integrated model;

calculating, using the first integrated model and in combination with the data
weight values, a dot product of the sample user vector and the sample item vector to
obtain a first predicted value about the historical user clicking the historical item, and

obtaining a first predicted probability based on the first predicted value;

iterative executing: according to a first difference between the first predicted probability and the historical probability, allocating model weight values for the at least two sub-models in the first integrated model based on the first difference, and
5 allocating first data weight values for the historical user and the historical item corresponding to the first predicted probability based on the first difference; and

adding one sub-model to the first integrated model to constitute a second integrated model to substitute the first integrated model, using the second integrated model as the first integrated model, calculating, using the first integrated model and in
10 combination with the first data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtaining a first predicted probability based on the first predicted value until satisfying a preset convergence condition.

In another embodiment based on the method of the present application, the
15 satisfying a preset convergence condition includes:

the first difference between the first predicted probability and the historical probability is less than a preset difference.

According to another aspect of the embodiments of the present disclosure, provided is an apparatus for deep learning-based recommendation, including: an
20 information obtaining unit, configured to respectively obtain related information of a target user and related information of a target item; a probability obtaining unit, configured to respectively use at least two sub-models in an integrated model to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-
25 models, the operating probability being used for indicating a probability, obtained based on one sub-model, about the target user operating the target item, and the integrated model including at least two sub-models; a probability synthesis unit, configured to obtain, based on the operating probabilities corresponding to the at least two sub-models, a target probability about the target user operating the target item;
30 and a recommending unit, configured to recommend the target item to the target user based on the target probability.

In another embodiment based on the apparatus of the present application, the related information of the target user includes: history information about a historical

item being operated by the target user, the historical item including at least one item operated by the target user;

and/or

the related information of the target item includes history information about
5 the target item being operated by a historical user, the historical user including at least one user who operated the target item.

In another embodiment based on the apparatus of the present application, the probability obtaining unit includes:

a vectorization module, configured to use the sub-models to respectively
10 vectorize the related information of the target user and the related information of the target item, thereby obtaining user vectors and item vectors; and

an operating probability module, configured to obtain the operating probabilities based on the user vectors and the item vectors.

In another embodiment based on the apparatus of the present application, the
15 operating probability module is configured to calculate dot products of the user vectors and the item vectors to obtain operate values, and obtaining the operating probabilities based on the operate values.

In another embodiment based on the apparatus of the present application, the probability synthesis unit is configured to respectively calculate products of the
20 operating probabilities and model weight values of the sub-models corresponding to the operating probabilities, each model weight value corresponding to one sub-model, and add the products corresponding to the at least two operating probabilities to obtain the target probability about the target user operating the target item.

In another embodiment based on the apparatus of the present application, the
25 recommending unit is configured to:

in response to the target probability being greater than or equal to a preset threshold, recommend the target item corresponding to the target probability to the target user; and

in response to the target probability being less than a preset threshold,
30 recommend no target item corresponding to the target probability to the target user.

In another embodiment based on the apparatus of the present application, the integrated model is obtained by training a relation between the at least two sub-models.

In another embodiment based on the apparatus of the present application, the apparatus further includes:

a training unit, configured to train the integrated model based on a historical user and a historical item, where a historical probability about the historical user operating the historical item is known.

In another embodiment based on the apparatus of the present application, the training unit includes:

a vector module, configured to establish a historical vector corresponding to the historical user and the historical item based on the historical user and the historical item;

a decomposing module, configured to decompose the historical vector using the integrated model to obtain at least one sample user vector and at least one sample item vector;

a probability predicting module, configured to obtain a predicted probability based on the sample user vector and the sample item vector, the predicted probability indicating the probability about the historical user operating the historical item predicted by the integrated model; and

a model training module, configured to train the integrated model based on the predicted probability and the historical probability.

In another embodiment based on the apparatus of the present application, the decomposing module is configured to decompose the historical vector using an initial integrated model to obtain at least one sample user vector and at least one sample item vector, the initial integrated model including one sub-model.

In another embodiment based on the apparatus of the present application, the probability predicting module is configured to calculate a dot product of the sample user vector and the sample item vector using the initial integrated model to obtain a predicted value about the historical user clicking the historical item, and obtain the predicted probability based on the predicted value.

In another embodiment based on the apparatus of the present application, the model training module includes:

a weight allocation module, configured to allocate, according to a difference between the predicted probability and the historical probability, data weight values for the historical user and the historical item corresponding to the predicted probability,

where the greater the difference is, the greater the data weight values are; and
an iterative training module, configured to iteratively train the integrated model based on the sample user vector and the sample item vector combined with the data weight values.

5 In another embodiment based on the apparatus of the present application, the iterative training module is configured to:

add one sub-model to the initial integrated model to constitute a first integrated model;

10 calculate, using the first integrated model and in combination with the data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtain a first predicted probability based on the first predicted value;

iterative execution: according to a first difference between the first predicted probability and the historical probability, allocate model weight values for the at least
15 two sub-models in the first integrated model based on the first difference, and allocate first data weight values for the historical user and the historical item corresponding to the first predicted probability based on the first difference;

add one sub-model to the first integrated model to constitute a second integrated model to substitute the first integrated model, use the second integrated
20 model as the first integrated model, calculate, using the first integrated model and in combination with the first data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtain a first predicted probability based on the first predicted value until satisfying a preset convergence condition.

25 In another embodiment based on the apparatus of the present application, the satisfying a preset convergence condition includes:

the first difference between the first predicted probability and the historical probability is less than a preset difference.

30 According to another aspect of the embodiments of the present disclosure, provided is an electronic device, including a processor, where the processor includes the apparatus for deep learning-based recommendation.

According to another aspect of embodiments of the present disclosure, provided is an electronic device, including a memory configured to store executable

instructions; and a processor configured to communicate with the memory to execute the executable instructions to complete operations of the method for deep learning-based recommendation as stated above.

According to another aspect of the embodiments of the present disclosure,
5 provided is a computer storage medium, having stored thereon instructions that, when executed by a processor, cause the processor to perform operations of the method for deep learning-based recommendation as stated above.

According to another aspect of the embodiments of the present disclosure,
provided is a computer program, including computer-readable codes, where when the
10 computer-readable codes run in a device, a processor in the device executes instructions for implementing the method for deep learning-based recommendation as stated above.

Based on the method and apparatus deep learning-based recommendation,
electronic device, medium, and program provided by the embodiments of the present
15 disclosure, related information of a target user and related information of a target item are respectively obtained; at least two sub-models in an integrated model are respectively used to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models; the operating probabilities are respectively obtained by means
20 of the integrated at least two sub-models, then a target probability about the target user operating the target item is obtained based on the at least two operating probabilities, and by means of the complementarity between the at least two sub-models, different data distributions that the at least two sub-models specialize are complemented, and the obtained probability value may more accurately predict the
25 probability about the target user clicking the target item; and the target item is recommended to the target user based on the target probability.

The following further describes in detail the technical solutions of the present disclosure with reference to the accompanying drawings and embodiments.

30

BRIEF DESCRIPTION OF DRAWINGS

The accompanying drawings constituting a part of the specification describe the embodiments of the present disclosure and are intended to explain the principles

of the present disclosure together with the descriptions.

With reference to the drawings, according to the detailed description, the present disclosure can be understood more clearly, where:

FIG. 1 is a flowchart of one embodiment of a method for deep learning-based recommendation according to the present disclosure;

FIG. 2 is a schematic diagram of one example of a method for deep learning-based recommendation according to the present disclosure;

FIG. 3 is a schematic diagram of one example of training an integrated model in a method for deep learning-based recommendation according to the present disclosure;

FIG. 4 is a schematic structural diagram of one embodiment of an apparatus for deep learning-based recommendation according to the present disclosure;

FIG. 5 is a schematic structural diagram of an electronic device for implementing a terminal device or a server in embodiments of the present disclosure.

DETAILED DESCRIPTION

Each exemplary embodiments of the present disclosure are described in detail with reference to the accompany drawings now. It should be noted that, unless otherwise stated, relative arrangement of the components and operations, the numerical expressions, and the values set forth in the embodiments are not intended to limit the scope of the present disclosure.

In addition, it should be understood that, for ease of description, the size of each part shown in the accompanying drawings is not drawn in actual proportion.

The following descriptions of at least one exemplary embodiment are merely illustrative actually, and are not intended to limit the present disclosure and the applications or uses thereof.

Technologies, methods and devices known to a person of ordinary skill in the related art may not be discussed in detail, but such technologies, methods and devices should be considered as a part of the specification in appropriate situations.

It should be noted that similar reference numerals and letters in the following accompanying drawings represent similar items. Therefore, once an item is defined in an accompanying drawing, the item does not need to be further discussed in the

subsequent accompanying drawings.

The embodiments of the present disclosure may be applied to electronic devices such as terminal devices, computer systems, and servers, which may operate with numerous other general-purpose or special-purpose computing system environments or configurations. Examples of well-known terminal devices, computing systems, environments, and/or configurations suitable for use together with the electronic devices such as terminal devices, computer systems, and servers include, but are not limited to, personal computer systems, server computer systems, thin clients, thick clients, handheld or laptop devices, microprocessor-based systems, set top boxes, programmable consumer electronics, network personal computers, small computer systems, large computer systems, distributed cloud computing environments that include any one of the systems, and the like.

The electronic devices such as terminal devices, computer systems, and servers may be described in the general context of computer system executable instructions (such as, program modules) executed by the computer systems. Generally, the program modules may include routines, programs, target programs, assemblies, logics, data structures, and the like, to perform specific tasks or implement specific abstract data types. The computer systems/servers may be practiced in the distributed cloud computing environments in which tasks are executed by remote processing devices that are linked through a communications network. In the distributed computing environments, program modules may be located in local or remote computing system storage media including storage devices.

FIG. 1 is a flowchart of one embodiment of a method for deep learning-based recommendation according to the present disclosure. As shown in FIG. 1, the method of this embodiment includes the following operations.

In operation 101, related information of a target user and related information of a target item are obtained.

The related information includes one or more of information related to an operation of the user, or information related to an operated item.

In one or more embodiments, the related information of the target user may include, but is not limited to: history information about a historical item being operated by the target user, the historical item including at least one item operated by the target user; the related information of the target item may include, but is not

limited to: history information about the target item being operated by a historical user, the historical user including at least one user who operated the target item. According to this embodiment, an item which a user is interested in may be recommended to the user by merely using historical operation information of the user
5 and historical information of the item being operated, without a need for acquiring attribute information of the user and the item, and the like, and thus the information processing process is simplified, and the recommendation speed is improved.

According to one or more embodiments of the present disclosure, in a certain website (for example, a video website or a shopping website), when there is a need for
10 recommending an item (for example, a commodity or a video) to a target user, there may be two circumstances: on the one hand, the target user may be an old user of the website (who carried out operations of clicking some items in the website before), at this time, a target item recommended to the target user may be an item which was operated by the target user according to types of items, shopping habits of the user,
15 and the like, or an item which is not operated by the target user, and at this time, historical operation information of the target user is required to be obtained; on the other hand, the target user may be a new user of the website (who does not carry out any operation of clicking an item in the website before), an item which was operated by another user in the website may be recommended to the target user, and at this
20 time, historical operation information of the target item is required to be obtained.

In an example, the operation 101 may be executed by a processor by invoking a corresponding instruction stored in a memory, or may be executed by an information obtaining unit 41 run by the processor.

In operation 102, at least two sub-models in an integrated model are
25 respectively used to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models.

The operating probability is used for indicating a probability, obtained based on one sub-model, about the target user operating the target item, and the integrated
30 model includes at least two sub-models; the integrated model is obtained by training, the training on the integrated model is based on a relation between the at least two sub-models, and the operating probabilities can respectively be obtained by means of the at least two trained sub-models.

In an example, the operation 102 may be executed by a processor by invoking a corresponding instruction stored in a memory, or may be executed by a probability obtaining unit 42 run by the processor.

5 In operation 103, a target probability about the target user operating the target item is obtained based on the operating probabilities corresponding to the at least two sub-models.

According to one or more embodiments of the present disclosure, based on a relation between the at least two trained sub-models, for example, weights corresponding to the at least two sub-models, the operating probabilities obtained
10 based on the at least two sub-models are combined in combination with the weights to obtain the target probability; and based on integration of multiple sub-models, accuracy of the integrated model is enhanced.

In an example, the operation 103 may be executed by a processor by invoking a corresponding instruction stored in a memory, or may be executed by a probability
15 synthesis unit 43 run by the processor.

In operation 104, the target item is recommended to the target user based on the target probability.

In an example, the operation 104 may be executed by a processor by invoking a corresponding instruction stored in a memory, or may be executed by a
20 recommending unit 44 run by the processor.

Based on the method for deep learning-based recommendation provided by the embodiment above of the present disclosure, a target user and a target item are received to respectively obtain related information of the target user and related information of the target item; at least two sub-models in an integrated model are
25 respectively used to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models; the operating probabilities are respectively obtained by means of the integrated at least two sub-models, then a target probability about the target user operating the target item is obtained based on the at least two operating
30 probabilities, and by means of the complementarity between the at least two sub-models, different data distributions that the at least two sub-models specialize are complemented, and the obtained probability value may more accurately predict the probability about the target user clicking the target item; and the target item is

recommended to the target user based on the target probability.

In another embodiment based on the method for deep learning-based recommendation of the present disclosure, based on at least one embodiment as stated above, the operation 102 includes:

5 using the sub-models to respectively vectorize the related information of the target user and the related information of the target item, thereby obtaining user vectors and item vectors; and

 obtaining the operating probabilities based on the user vectors and the item vectors. According to one or more embodiments of the present disclosure, dot products of the user vectors and the item vectors are calculated to obtain operate values, and the operating probabilities are obtained based on the operate values.

 According to one or more embodiments of the present disclosure, the sub-models in this embodiment may be implemented by means of Weighted Matrix Factorization (WMF). In order to achieve probability calculation, the related information of the target user and the related information of the target item are respectively vectored by means of the trained sub-model, the dimensionality of vectorization is decided by means of the trained sub-models, and the user vectors and the item vectors after vectorization satisfy the requirement for executing the dot product operation; by calculating the dot products of the user vectors and the item vectors, the operate values are obtained, and at this time, the operate values are numerical values; in order to obtain the probability values, the operate values may be subjected to processing such as normalization so as to obtain the operating probabilities.

 In another embodiment based on the method for deep learning-based recommendation of the present disclosure, based on at least one embodiment as stated above, the operation 103 includes:

 respectively calculating products of the operating probabilities and model weight values of the sub-models corresponding to the operating probabilities, each model weight value corresponding to one sub-model; and

30 adding the products corresponding to the at least two operating probabilities to obtain the target probability about the target user operating the target item.

 According to one or more embodiments of the present disclosure, an appropriate model weight value is allocated to each sub-mode by means of training,

the proportion of the operating probability obtained by each sub-model in the total target probability is controlled by means of the model weight value, and the higher the weight value is, the greater effect the sub-model may have on item recommendation.

In one or more embodiments, the operation 104 may include:

5 in response to the target probability being greater than or equal to a preset threshold, recommending the target item corresponding to the target probability to the target user; and

 in response to the target probability being less than a preset threshold, recommending no target item corresponding to the target probability to the target user.

10 In the method for deep learning-based recommendation in this embodiment, the final purpose is to recommend to a user a target item which the target user may be more interested in; therefore, when the target probability is greater than or equal to the preset threshold, it indicates that the probability about the target user clicking the target item is greater than or equal to the preset threshold, and recommending the
15 target item to the target user may effectively improve the operation probability of user, thereby achieving the purpose of effective recommendation.

 FIG. 2 is a schematic diagram of one example of a method for deep learning-based recommendation according to the present disclosure. As shown in FIG. 2, each of users corresponds to a vector representation, i.e., an embedding, in each sub-model, and similarly, each of the items also corresponds to an embedding in each sub-model.
20 An inner product of vectors corresponding to a pair of a user and an item is a predicted probability (i.e., a prediction) provided by a current sub-model, which is represented as $P(r|u_1, v_1)$, $P(r|u_2, v_2)$, ..., $P(r|u_n, v_n)$ in the drawing; a final target probability (final prediction) can be obtained by means of weight addition of the probabilities of the multiple sub-models; in the drawing, α_1 , α_2 , ..., α_n respectively
25 represent the model weight values corresponding to each sub-model, and the target probability is obtained by weighted addition in combination with the predicted probabilities obtained by the sub-models and the model weight values.

 In another embodiment based on the method for deep learning-based recommendation of the present disclosure, based on at least one embodiment as stated
30 above, the method further includes:

 training the integrated model based on a historical user and a historical item,

 where a historical probability about the historical user operating the historical

item is known.

In one or more embodiments, in order to obtain a more accurate target probability, before the integrated model is applied, the integrated model is required to be trained; information used in training is historical users and historical items based on known operation information, and by analyzing the operation information of the historical users and the historical items, one or more newly added items may be recommended to any user in the historical users, and/or any of one or more historical items may be recommended to one or more new users.

In an example of the at least one embodiment above of the method for deep learning-based recommendation of the present disclosure, the training the integrated model based on a historical user and a historical item includes:

establishing a historical vector corresponding to the historical user and the historical item based on the historical user and the historical item;

decomposing the historical vector using the integrated model to obtain at least one sample user vector and at least one sample item vector;

obtaining a predicted probability based on the sample user vector and the sample item vector, the predicted probability indicating the probability about the historical user operating the historical item predicted by the integrated model; and

training the integrated model based on the predicted probability and the historical probability.

In this embodiment, because the operation information of the historical user and the historical item is known, a matrix vector may be established based on which historical item the historical operated and based on which historical user the historical item is operated by; by taking the historical user as the abscissa (or the ordinate) and taking the historical item as the ordinate (or the abscissa), a value exists between the historical user and the historical item which have an operation relation, and the magnitude of the value is decided by the operation information. The value between the historical user and the historical item which do not have any operation relation may be set as 0.

In an example of the at least one embodiments above of the method for deep learning-based recommendation of the present disclosure, the decomposing the historical vector using the integrated model to obtain at least one sample user vector and at least one sample item vector, and obtaining a predicted probability based on the

sample user vector and the sample item vector includes:

decomposing the historical vector using an initial integrated model to obtain at least one sample user vector and at least one sample item vector, the initial integrated model including one sub-model; and

- 5 calculating a dot product of the sample user vector and the sample item vector using the initial integrated model to obtain a predicted value about the historical user clicking the historical item, and obtaining the predicted probability based on the predicted value.

According to one or more embodiments of the present disclosure, the process
10 of training the integrated model is training the sub-models one by one, and the initial integrated model merely includes one sub-model; the sub-model is trained, and different data weight values are allocated, based on the obtained predicted probability and the known historical probability, to data for training the integrated model; a new integrated model obtained after one sub-model is added is trained according to the
15 historical users and the historical items to which different data weight values are allocated, and then new data weight values are allocated based on the new predicted probability, and so forth in an iterative manner; data weight values are reallocated to the historical users and the historical items each time until convergence is achieved. Moreover, there is no need to restrict the number of iterations in advance in this
20 embodiment, and the training can be carried out until convergence of the integrated model is achieved, or computing resources are used out. Usually, a relatively good effect can be obtained after the first few iterations in this embodiment.

FIG. 3 is a schematic diagram of one example of training an integrated model in a method for deep learning-based recommendation according to the present
25 disclosure. As shown in FIG. 3, input in this example is operation information of all the historical users and all the historical items, so as to obtain historical vectors (represented as R in FIG. 3); output is vectors (embedding) (represented as U and V in FIG. 3) obtained by decomposition of the integrated model on the users and the items, and a final predicted probability (represented as P in FIG. 3, including: $P_0, P_1, \dots, P_n$);
30 in each turn of iterative training, the probabilities given by the integrated model are compared with an original evaluation matrix R , a higher weight is given to the predicted probability with a greater error, and a new sub-model is trained by using the obtained training weight, i.e., new U_i and V_i ; and then the probability given by the

new sub-model is added to the integrated P, and so forth in an iterative manner.

In one or more embodiments, the training the integrated model based on the predicted probability and the historical probability includes:

allocating, according to a difference between the predicted probability and the
5 historical probability, data weight values for the historical user and the historical item
corresponding to the predicted probability, where the greater the difference is, the
greater the data weight values are; and

iteratively training the integrated model based on the sample user vector and
the sample item vector combined with the data weight values.

10 The integrated model is trained based on the data weight value reallocated
each time in combination with the sample user vector and the sample item vector, and
in each allocation of the data weight value, the historical user and/or historical item
having a greater difference is allocated with a greater weight value.

In one or more embodiments, the process of iteratively training the integrated
15 model includes:

adding one sub-model to the initial integrated model to constitute a first
integrated model;

calculating, using the first integrated model and in combination with the data
weight values, a dot product of the sample user vector and the sample item vector to
20 obtain a first predicted value about the historical user clicking the historical item, and
obtaining a first predicted probability based on the first predicted value

iterative execution: according to a first difference between the first predicted
probability and the historical probability, allocating model weight values for the at
least two sub-models in the first integrated model based on the first difference, and
25 allocating first data weight values for the historical user and the historical item
corresponding to the first predicted probability based on the first difference; and

adding one sub-model to the first integrated model to constitute a second
integrated model to substitute the first integrated model, using the second integrated
model as the first integrated model, calculating, using the first integrated model and in
30 combination with the first data weight values, a dot product of the sample user vector
and the sample item vector to obtain a first predicted value about the historical user
clicking the historical item, and obtaining a first predicted probability based on the
first predicted value until satisfying a preset convergence condition.

According to one or more embodiments of the present disclosure, in the process of iteratively training the integrated model, the process of each iteration may be represented by means of formula (1):

$$p_{\alpha}(r|i,j) = (1 - \alpha)p(r|i,j) + \alpha p'(r|i,j) \quad \text{Formula (1)}$$

5 where $p_{\alpha}(r|i,j)$ is the predicted probability outputted by the integrated model after the latest combination; $p_{\alpha}(r|i,j)$ is the predicted probability outputted by the current integrated model (the probability that the number of times user i operating item j is r); $(1 - \alpha)$ is the model weight corresponding to the current integrated model; $p'(r|i,j)$ is the predicted probability outputted by the newly added sub-model; α is the model weight corresponding to the newly added sub-model. For example, when the current integrated model is the initial integrated model, the integrated model after the latest combination is a first integrated model; and when the current integrated model is the first integrated model, the integrated model after the latest combination is a second integrated model.

15 In each iterative training, in addition to adjusting the model weight of the sub-model, the data weight corresponding to the historical user and the historical item is further adjusted according to the difference between the predicted probability and the historical probability, and the calculation formula for the data weight is as shown in formula (2):

$$20 \quad \rho_{ij} = \frac{\alpha \bar{P}}{\alpha \bar{P} + (1 - \alpha) P(r|i,j)} \quad \text{Formula (2)}$$

where ρ_{ij} is the data weight; $\bar{P} = \iint p'(r|u,v) p_0(u) p_0(v) du dv$ represents the integration of $p'(r|u,v)$, where i and j are serial numbers, indicating the i -th user and the j -th item, u is the vector corresponding to the user, v is the vector corresponding to the item, and according to the formula of integration of $\bar{P}$, it can be derived that $\bar{P} = p'(r|i,j)$. Therefore, formula (2) may be converted in to formula (3):

$$25 \quad \rho_{ij} = \frac{1}{(1 + \gamma p(r|i,j))} \quad \text{Formula (3)}$$

where $\gamma = \frac{1 - \alpha}{\alpha \bar{P}}$ is a set parameter. Therefore, the magnitude of the data weight is decided based on the current integrated model, and the data weight changes

as the integrated model changes.

Optionally, the satisfying a preset convergence condition includes:

the first difference between the first predicted probability and the historical probability obtained in the iterative process is less than a preset difference. The preset
5 convergence condition may further include that all the calculation resources are used out.

Any deep learning-based recommendation method provided by the embodiments of the present application is executed by any appropriate device having data processing capability, including, but not limited to, a terminal device and a
10 server. Alternatively, any deep learning-based recommendation method provided in the embodiments of the present application is executed by a processor, for example, any image processing method mentioned in the embodiments of the present application is executed by the processor by invoking a corresponding instruction stored in a memory. Details are not described below again.

15 A person of ordinary skill in the art may understand that all or some operations for implementing the foregoing method embodiments are achieved by a program by instructing related hardware; the foregoing program can be stored in a computer-readable storage medium; when the program is executed, operations including the foregoing method embodiments are executed. Moreover, the foregoing storage
20 medium includes various media capable of storing a program code such as an ROM, an RAM, a magnetic disk, or an optical disk.

FIG. 4 is a schematic structural diagram of one embodiment of a deep learning-based recommendation apparatus according to the present application. The apparatus of this embodiment is used for implementing at least one of the foregoing
25 embodiments of the method according to the present application. As shown in FIG. 4, the apparatus of this embodiment includes:

an information obtaining unit 41, configured to respectively obtain related information of a target user and related information of a target item.

The related information includes one or more pieces of information used for
30 describing information related to an operation of the user, or information related to an operated item.

In one or more embodiments, the related information of the target user may include, but is not limited to: history information about a historical item being

operated by the target user, the historical item including at least one item operated by the target user; the related information of the target item may include, but is not limited to: history information about the target item being operated by a historical user, the historical user including at least one user who operated the target item.

5 According to this embodiment, an item which a user is interested in may be recommended to the user by merely using historical operation information of the user and historical information of the item being operated, without a need for acquiring attribute information of the user and the item, and the like, and thus the information processing process is simplified, and the recommendation speed is improved.

10 A probability obtaining unit 42, configured to respectively use at least two sub-models in an integrated model to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models,

where the operating probability is used for indicating a probability, obtained
15 based on one sub-model, about the target user operating the target item, and the integrated model includes at least two sub-models; the integrated model is obtained by training a relation between the at least two sub-models;

a probability synthesis unit 43, configured to obtain, based on the operating probabilities corresponding to the at least two sub-models, a target probability about
20 the target user operating the target item; and

a recommending unit 44, configured to recommend the target item to the target user based on the target probability.

Based on the apparatus for deep learning-based recommendation provided by the embodiment above of the present disclosure, a target user and a target item are
25 received to respectively obtain related information of the target user and related information of the target item; at least two sub-models in an integrated model are respectively used to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models; the operating probabilities are respectively obtained by means
30 of the integrated at least two sub-models, then a target probability about the target user operating the target item is obtained based on the at least two operating probabilities, and by means of the complementarity between the at least two sub-models, different data distributions that the at least two sub-models specialize are

complemented, and the obtained probability value may more accurately predict the probability about the target user clicking the target item; and the target item is recommended to the target user based on the target probability.

In another embodiment based on the apparatus for deep learning-based recommendation of the present disclosure, based on at least one embodiment as stated above, the probability obtaining unit 42 includes:

a vectorization module, configured to use the sub-models to respectively vectorize the related information of the target user and the related information of the target item, thereby obtaining user vectors and item vectors; and

an operating probability module, configured to obtain the operating probabilities based on the user vectors and the item vectors.

According to one or more embodiments of the present disclosure, the sub-models in this embodiment may be implemented by means of Weighted Matrix Factorization (WMF). In order to achieve probability calculation, the related information of the target user and the related information of the target item are respectively vectored by means of the trained sub-model, the dimensionality of vectorization is decided by means of the trained sub-models, and the user vectors and the item vectors after vectorization satisfy the requirement for executing the dot product operation; by calculating the dot products of the user vectors and the item vectors, the operate values are obtained, and at this time, the operate values are numerical values; in order to obtain the probability values, the operate values may be subjected to processing such as normalization so as to obtain the operating probabilities.

According to one or more embodiments of the present disclosure, the operating probability module is configured to calculate dot products of the user vectors and the item vectors to obtain operate values, and obtain the operating probabilities based on the operate values.

In another embodiment based on the apparatus for deep learning-based recommendation of the present disclosure, based on at least one embodiment as stated above, the probability synthesis unit 43 is configured to respectively calculate products of the operating probabilities and model weight values of the sub-models corresponding to the operating probabilities, each model weight value corresponding to one sub-model; and add the products corresponding to the at least two operating

probabilities to obtain the target probability about the target user operating the target item.

According to one or more embodiments of the present disclosure, an appropriate model weight value is allocated to each sub-mode by means of training,
5 the proportion of the operating probability obtained by each sub-model in the total target probability is controlled by means of the model weight value, and the higher the weight is, the greater effect the sub-model may have on item recommendation.

In one or more embodiments, the recommending unit 44 is configured to:

in response to the target probability being greater than or equal to a preset
10 threshold, recommend the target item corresponding to the target probability to the target user; and

in response to the target probability being less than a preset threshold, recommend no target item corresponding to the target probability to the target user.

In another embodiment based on the apparatus for deep learning-based
15 recommendation of the present disclosure, based on at least one embodiment as stated above, the apparatus further includes:

a training unit, configured to train the integrated model based on a historical user and a historical item,

where a historical probability about the historical user operating the historical
20 item is known.

In one or more embodiments, in order to obtain a more accurate target probability, before the integrated model is applied, the integrated model is required to be trained; information used in training is historical users and historical items based on known operation information, and by analyzing the operation information of the
25 historical users and the historical items, one or more newly added items may be recommended to any user in the historical users, and/or any of one or more historical items may be recommended to one or more new users.

In an example of the at least one embodiment above of the apparatus for deep learning-based recommendation of the present disclosure, the training unit includes:

30 a vector module, configured to establish a historical vector corresponding to the historical user and the historical item based on the historical user and the historical item;

a decomposing module, configured to decompose the historical vector using

the integrated model to obtain at least one sample user vector and at least one sample item vector;

probability predicting module, configured to obtain a predicted probability based on the sample user vector and the sample item vector, the predicted probability
5 indicating the probability about the historical user operating the historical item predicted by the integrated model; and

a model training module, configured to train the integrated model based on the predicted probability and the historical probability.

In this embodiment, because the operation information of the historical user
10 and the historical item is known, a matrix vector may be established based on which historical item the historical operated and based on which historical user the historical item is operated by; by taking the historical user as the abscissa (or the ordinate) and taking the historical item as the ordinate (or the abscissa), a value exists between the historical user and the historical item which have an operation relation, and the
15 magnitude of the value is decided by the operation information. The value between the historical user and the historical item which do not have any operation relation may be set as 0.

In an example of the at least one embodiments above of the apparatus for deep learning-based recommendation of the present disclosure, the decomposing module is
20 configured to decompose the historical vector using an initial integrated model to obtain at least one sample user vector and at least one sample item vector, the initial integrated model including one sub-model.

In an example of the at least one embodiments above of the apparatus for deep learning-based recommendation of the present disclosure, the probability predicting
25 module is configured to calculate a dot product of the sample user vector and the sample item vector using the initial integrated model to obtain a predicted value about the historical user clicking the historical item, and obtain the predicted probability based on the predicted value.

In an example of the at least one embodiment above of the apparatus for deep
30 learning-based recommendation of the present disclosure, the model training module includes:

a weight allocation module, configured to allocate, according to a difference between the predicted probability and the historical probability, data weight values for

the historical user and the historical item corresponding to the predicted probability, where the greater the difference is, the greater the data weight values are; and

an iterative training module, configured to iteratively train the integrated model based on the sample user vector and the sample item vector combined with the data weight values.

In an example of the at least one embodiment above of the apparatus for deep learning-based recommendation of the present disclosure, the iterative training module is configured to:

add one sub-model to the initial integrated model to constitute a first integrated model;

calculate, using the first integrated model and in combination with the data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtain a first predicted probability based on the first predicted value

iterative execution: according to a first difference between the first predicted probability and the historical probability, allocate model weight values for the at least two sub-models in the first integrated model based on the first difference, and allocate first data weight values for the historical user and the historical item corresponding to the first predicted probability based on the first difference; and

add one sub-model to the first integrated model to constitute a second integrated model to substitute the first integrated model, use the second integrated model as the first integrated model, calculate, using the first integrated model and in combination with the first data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtain a first predicted probability based on the first predicted value until satisfying a preset convergence condition.

In one or more embodiments, the satisfying a preset convergence condition includes:

the first difference between the first predicted probability and the historical probability obtained in the iterative process is less than a preset difference.

According to another aspect of the embodiments of the present disclosure, provided is an electronic device, including a processor, where the processor includes the apparatus for deep learning-based recommendation.

An electronic device provided according to one aspect of the embodiments of the present disclosure includes: a memory configured to store an executable instruction; and

5 a processor, configured to communicate with the memory to execute the executable instruction to achieve operations of any embodiment of the method for deep learning-based recommendation of the present disclosure

A computer storage medium provided according to one aspect of the embodiments of the present disclosure is configured to store a computer readable instruction, where when the instruction is executed, the operations according to any
10 one of the foregoing embodiments of the method for deep learning-based recommendation of the present disclosure are implemented.

A computer program provided according to one aspect of the embodiments of the present disclosure includes a computer readable code, where when the computer readable code runs on a device, a processor in the device executes an instruction for
15 implementing any of the foregoing embodiments of the method for deep learning-based recommendation of the present disclosure.

The embodiments of the present disclosure further provide an electronic device which, for example, may be a mobile terminal, a Personal Computer (PC), a tablet computer, a server, and the like. Referring to FIG. 5 below, a schematic
20 structural diagram of an electronic device 500, which may be a terminal device or a server, suitable for implementing an embodiment of the present disclosure is shown. As shown in FIG. 5, the computer system 500 includes one or more processors, a communication section, and the like. The one or more processors are, for example, one or more Central Processing Units (CPUs) 501 and/or one or more Graphic
25 Processing Units (GPUs) 513, and the processors may execute appropriate actions and processing according to executable instructions stored in a Read-Only Memory (ROM) 502 or executable instructions loaded from a storage part 508 to a Random Access Memory (RAM) 503. The communication part 512 may include, but is not limited to, a network card. The network card may include, but is not limited to, an
30 Infiniband (IB) network card.

The processor may communicate with the ROM 502 and/or the RAM 503 to execute executable instructions, is connected to the communication part 512 by means of a bus 504, and communicates with other target devices by means of the

communication part 512, so as to complete corresponding operations of any of the methods provided by the embodiments of the present disclosure, for example, respectively obtaining related information of a target user and related information of a target item; respectively using at least two sub-models in an integrated model to
5 obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models; obtaining, based on the operating probabilities corresponding to the at least two sub-models, a target probability about the target user operating the target item; and recommending the target item to the target user based on the target probability.

10 In addition, the RAM 503 may further store various programs and data required for operations of an apparatus. The CPU 501, the ROM 502, and the RAM 503 are connected to each other via the bus 504. In the presence of the RAM 503, the ROM 502 is an optional module. The RAM 503 stores executable instructions, or writes the executable instructions into the ROM 502 during running, where the
15 executable instructions cause the processor 501 to execute corresponding operations of the foregoing method for deep learning-based recommendation. An Input/Output (I/O) interface 505 is also connected to the bus 504. The communication part 512 is integrated, or is configured to have multiple sub-modules (for example, multiple IB network cards) connected to the bus.

20 The following components are connected to the I/O interface 505: an input section 506 including a keyboard, a mouse and the like; an output section 507 including a Cathode-Ray Tube (CRT), a Liquid Crystal Display (LCD), a speaker and the like; the storage section 508 including a hard disk and the like; and a communication section 509 of a network interface card including an LAN card, a
25 modem and the like. The communication section 509 performs communication processing via a network such as the Internet. A drive 510 is also connected to the I/O interface 505 according to requirements. A removable medium 511 such as a magnetic disk, an optical disk, a magneto-optical disk, a semiconductor memory or the like is mounted on the drive 510 according to requirements, so that a computer program read
30 from the removable medium is installed on the storage section 508 according to requirements.

It should be noted that the architecture illustrated in FIG. 5 is merely an optional implementation mode. During specific practice, the number and types of the

components in FIG. 5 may be selected, decreased, increased, or replaced according to actual requirements. Different functional components may be separated or integrated or the like. For example, the GPU and the CPU may be separated, or the GPU may be integrated on the CPU, and the communication portion may be separated from or
5 integrated on the CPU or the GPU or the like. These alternative implementations all fall within the scope of protection of the present disclosure.

Particularly, a process described above with reference to a flowchart according to the embodiments of the present disclosure is implemented as a computer software program. For example, embodiments of the present disclosure include a computer
10 program product. The computer program product includes a computer program tangibly included in a machine-readable medium. The computer program includes a program code for performing a method shown in the flowchart. The program code may include instructions for performing steps of the method for deep learning-based recommendation provided in the embodiments of the present disclosure, for example,
15 respectively obtaining related information of a target user and related information of a target item; respectively using at least two sub-models in an integrated model to obtain, based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models; obtaining, based on the operating probabilities corresponding to the at least
20 two sub-models, a target probability about the target user operating the target item; and recommending the target item to the target user based on the target probability. In such embodiments, the computer program is downloaded and installed from the network through the communication section 509, and/or is installed from the removable medium 511. The computer program, when being executed by the CPU
25 501, executes the foregoing functions defined in the methods of the present disclosure.

The embodiments in the specification are all described in a progressive manner, for same or similar parts in the embodiments, refer to these embodiments, and each embodiment focuses on a difference from other embodiments. The system
30 embodiments correspond to the method embodiments substantially and therefore are only described briefly, and for the associated part, refer to the descriptions of the method embodiments.

The methods, apparatuses, and devices in the present disclosure are

implemented in many manners. For example, the methods, apparatuses, and devices in the present disclosure are implemented with software, hardware, firmware, or any combination of software, hardware, and firmware. The foregoing sequence of the operations of the method is merely for description, and unless otherwise stated particularly, the operations of the method in the present disclosure are not limited to the described sequence. In addition, in some embodiments, the present disclosure is also implemented as programs recorded in a recording medium. The programs include machine-readable instructions for implementing the methods according to the present disclosure. Therefore, the present disclosure further covers the recording medium storing the programs for performing the methods according to the present disclosure.

The descriptions of the present disclosure are provided for the purpose of examples and description, and are not intended to be exhaustive or limit the present disclosure to the disclosed form. Many modifications and changes are obvious to a person of ordinary skill in the art. The embodiments are selected and described to better describe a principle and an actual application of the present disclosure, and to make a person of ordinary skill in the art understand the present disclosure, so as to design various embodiments with various modifications applicable to particular use.

CLAIMS

1. A method for deep learning-based recommendation, comprising:

5 respectively obtaining related information of a target user and related information of a target item, the related information comprising one or more of information related to an operation of the user, or information related to an operated item;

 respectively using at least two sub-models in an integrated model to obtain,
10 based on the related information of the target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models, wherein the operating probability is used for indicating a probability, obtained based on one sub-model, about the target user operating the target item, and the integrated model comprises at least two sub-models;

15 obtaining, based on the operating probabilities corresponding to the at least two sub-models, a target probability about the target user operating the target item; and

 recommending the target item to the target user based on the target probability.

20 2. The method according to claim 1, wherein the related information of the target user comprises: history information about a historical item being operated by the target user, the historical item comprising at least one item operated by the target user;

 and/or

 the related information of the target item comprises history information about
25 the target item being operated by a historical user, the historical user comprising at least one user who operated the target item.

3. The method according to claim 1 or 2, wherein the respectively using at least two sub-models in an integrated model to obtain, based on the related information of the
30 target user and the related information of the target item, operating probabilities corresponding to the at least two sub-models comprises:

 using the sub-models to respectively vectorize the related information of the target user and the related information of the target item to obtain user vectors and

item vectors; and

obtaining the operating probabilities based on the user vectors and the item vectors.

- 5 4. The method according to claim 3, wherein the obtaining the operating probabilities based on the user vectors and the item vectors comprises:

calculating dot products of the user vectors and the item vectors to obtain operate values, and obtaining the operating probabilities based on the operate values.

- 10 5. The method according to any one of claims 1 to 4, wherein the obtaining, based on the operating probabilities corresponding to the at least two sub-models, a target probability about the target user operating the target item comprises:

respectively calculating products of the operating probabilities and model weight values of the sub-models corresponding to the operating probabilities, each
15 model weight value corresponding to one sub-model; and

adding the products corresponding to the at least two operating probabilities to obtain a target probability about the target user operating the target item.

- 20 6. The method according to any one of claims 1 to 5, wherein the recommending the target item to the target user based on the target probability comprises:

in response to the target probability being greater than or equal to a preset threshold, recommending the target item corresponding to the target probability to the target user; and

- 25 in response to the target probability being less than a preset threshold, recommending no target item corresponding to the target probability to the target user.

7. The method according to any one of claims 1 to 6, wherein the integrated model is obtained by training a relation between at least two sub-models.

- 30 8. The method according to claim 7, wherein a training mode for the integrated model comprises:

training the integrated model based on a historical user and a historical item,

wherein a historical probability about the historical user operating the historical item is known.

9. The method according to claim 8, wherein the training the integrated model based on a historical user and a historical item comprises:

establishing a historical vector corresponding to the historical user and the historical item based on the historical user and the historical item;

decomposing, using the integrated model, the historical vector to obtain at least one sample user vector and at least one sample item vector;

10 obtaining a predicted probability based on the sample user vector and the sample item vector, the predicted probability indicating a probability about the historical user operating the historical item predicted by the integrated model; and

training the integrated model based on the predicted probability and the historical probability.

15

10. The method according to claim 9, wherein the decomposing, using the integrated model, the historical vector to obtain at least one sample user vector and at least one sample item vector comprises:

20 decomposing, using an initial integrated model, the historical vector to obtain at least one sample user vector and at least one sample item vector, the initial integrated model comprising one sub-model.

11. The method according to claim 10, wherein the obtaining a predicted probability based on the sample user vector and the sample item vector comprises:

25 calculating, using the initial integrated model, a dot product of the sample user vector and the sample item vector to obtain a predicted value about the historical user clicking the historical item, and obtaining a predicted probability based on the predicted value.

30 12. The method according to any one of claims 9 to 11, wherein the training the integrated model based on the predicted probability and the historical probability comprises:

allocating, according to a difference between the predicted probability and the historical probability, data weight values for the historical user and the historical item corresponding to the predicted probability, wherein the greater the difference is, the greater the data weight values are; and

5 iteratively training the integrated model based on the sample user vector and the sample item vector combined with the data weight values.

13. The method according to claim 12, where the iteratively training the integrated model based on the sample user vector and the sample item vector combined with the
10 data weight values comprises:

 adding one sub-model to the initial integrated model to constitute a first integrated model;

 calculating, using the first integrated model and in combination with the data weight values, a dot product of the sample user vector and the sample item vector to
15 obtain a first predicted value about the historical user clicking the historical item, and obtaining a first predicted probability based on the first predicted value;

 iterative executing: according to a first difference between the first predicted probability and the historical probability, allocating model weight values for the at least two sub-models in the first integrated model based on the first difference, and
20 allocating first data weight values for the historical user and the historical item corresponding to the first predicted probability based on the first difference; and

 adding one sub-model to the first integrated model to constitute a second integrated model to substitute the first integrated model, using the second integrated model as the first integrated model, calculating, using the first integrated model and in
25 combination with the first data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtaining a first predicted probability based on the first predicted value until satisfying a preset convergence condition.

30 14. The method according to claim 13, wherein the satisfying a preset convergence condition comprises:

 the first difference between the first predicted probability and the historical probability is less than a preset difference.

15. An apparatus for deep learning-based recommendation, comprising:

an information obtaining unit, configured to respectively obtain related information of a target user and related information of a target item, the related information comprising one or more of information related to an operation of the user,
5 or information related to an operated item;

a probability obtaining unit, configured to respectively use at least two sub-models in an integrated model to obtain, based on the related information of the target user and the related information of the target item, operating probabilities
10 corresponding to the at least two sub-models, wherein the operating probability is used for indicating a probability, obtained based on one sub-model, about the target user operating the target item, and the integrated model comprises at least two sub-models;

a probability synthesis unit, configured to obtain, based on the operating probabilities corresponding to the at least two sub-models, a target probability about
15 the target user operating the target item; and

a recommending unit, configured to recommend the target item to the target user based on the target probability.

16. The apparatus according to claim 15, wherein the related information of the target user comprises: history information about a historical item being operated by the target user, the historical item comprising at least one item operated by the target user;
20 and/or

the related information of the target item comprises history information about
25 the target item being operated by a historical user, the historical user comprising at least one user who operated the target item.

17. The apparatus according to claim 15 or 16, wherein the probability obtaining unit comprises:

a vectorization module, configured to use the sub-models to respectively vectorize the related information of the target user and the related information of the target item to obtain user vectors and item vectors; and
30

an operating probability module, configured to obtain the operating probabilities based on the user vectors and the item vectors.

18. The apparatus according to claim 17, wherein the operating probability module is
5 configured to calculate dot products of the user vectors and the item vectors to obtain operate values, and obtain the operating probabilities based on the operate values.

19. The apparatus according to any one of claims 15 to 18, wherein the probability
synthesis unit is configured to
10 respectively calculate products of the operating probabilities and model weight
values of the sub-models corresponding to the operating probabilities, each model
weight value corresponding to one sub-model, and
 add the products corresponding to the at least two operating probabilities to
obtain the target probability about the target user operating the target item.

15
20. The apparatus according to any one of claims 15 to 19, wherein the
recommending unit is configured to:
 in response to the target probability being greater than or equal to a preset
threshold, recommend the target item corresponding to the target probability to the
20 target user; and
 in response to the target probability being less than a preset threshold,
recommend no target item corresponding to the target probability to the target user.

21. The apparatus according to any one of claims 15 to 20, wherein the integrated
25 model is obtained by training a relation between the at least two sub-models.

22. The apparatus according to claim 21, further comprising:
 a training unit, configured to train the integrated model based on a historical
user and a historical item, wherein a historical probability about the historical user
30 operating the historical item is known.

23. The apparatus according to claim 22, wherein the training unit comprises:

a vector module, configured to establish a historical vector corresponding to the historical user and the historical item based on the historical user and the historical item;

5 a decomposing module, configured to decompose, using the integrated model, the historical vector to obtain at least one sample user vector and at least one sample item vector;

a probability predicting module, configured to obtain a predicted probability based on the sample user vector and the sample item vector, the predicted probability indicating a probability about the historical user operating the historical item
10 predicted by the integrated model; and

a model training module, configured to train the integrated model based on the predicted probability and the historical probability.

24. The apparatus according to claim 23, wherein the decomposing module is
15 configured to:

decompose, using an initial integrated model, the historical vector to obtain at least one sample user vector and at least one sample item vector, the initial integrated model comprising one sub-model.

20 25. The apparatus according to claim 24, wherein the probability predicting module is configured to:

calculate, using the initial integrated model, a dot product of the sample user vector and the sample item vector to obtain a predicted value about the historical user clicking the historical item, and obtaining the predicted probability based on the
25 predicted value.

26. The apparatus according to any one of claims 23-25, wherein the model training module comprises:

a weight allocation module, configured to allocate, according to a difference
30 between the predicted probability and the historical probability, data weight values for the historical user and the historical item corresponding to the predicted probability, wherein the greater the difference is, the greater the data weight values are; and

an iterative training module, configured to iteratively train the integrated model based on the sample user vector and the sample item vector combined with the data weight values.

5 27. The apparatus according to claim 26, wherein the iterative training module is configured to:

add one sub-model to the initial integrated model to constitute a first integrated model;

10 calculate, using the first integrated model and in combination with the data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtain a first predicted probability based on the first predicted value;

iterative execute: according to a first difference between the first predicted probability and the historical probability, allocate corresponding model weight values
15 for the at least two sub-models in the first integrated model based on the first difference, and allocate first data weight values for the historical user and the historical item corresponding to the first predicted probability based on the first difference; and

add one sub-model to the first integrated model to constitute a second
20 integrated model to substitute the first integrated model, use the second integrated model as the first integrated model, calculate, using the first integrated model and in combination with the first data weight values, a dot product of the sample user vector and the sample item vector to obtain a first predicted value about the historical user clicking the historical item, and obtain a first predicted probability based on the first
25 predicted value until satisfying a preset convergence condition.

28. The apparatus according to claim 27, wherein the satisfying a preset convergence condition comprises:

30 the first difference between the first predicted probability and the historical probability is less than a preset difference.

29. An electronic device, comprising a processor, the processor comprising the

apparatus for deep learning-based recommendation according to any one of claims 15 to 28.

30. An electronic device, comprising: a memory, configured to store executable
5 instructions; and

a processor, configured to communicate with the memory to execute the executable instructions to complete operations of a method for deep learning-based recommendation according to any one of claims 1 to 14.

10 31. A computer storage medium for storing computer-readable instructions, wherein the instructions are executed to implement operations of a method for deep learning-based recommendation according to any one of claims 1 to 14.

32. A computer program, comprising computer-readable codes, wherein when the
15 computer-readable codes run in a device, a processor in the device executes instructions for implementing a method for deep learning-based recommendation according to any one of claims 1 to 14.