

METHODS AND APPARATUSES FOR TRAINING BEHAVIOR PREDICTION MODEL

TECHNICAL FIELD

[0001] One or more implementations of the present specification relate to the field of computer technologies, and more specifically, to methods and apparatuses for training a behavior prediction model in the field of computer technologies.

BACKGROUND

[0002] In daily production or life, users often perform various behaviors. Through data analysis techniques, people can perform analysis and processing on a historical behavior of a user, so as to predict a time at which a future behavior of the user occurs. For example, products that a user has previously clicked can be organized into a time sequence of click behaviors in the order of the time of the clicks. By analyzing the time sequence of the click behaviors, a time at which the user will click a product in the future can be predicted.

[0003] For a multi-behavior sequence of a user in an uneven continuous time domain, a conventional behavior sequence modeling method is to discretize the multi-behavior sequence into behaviors in multiple even time intervals, and time aggregation is performed on multiple behaviors in each time interval to obtain a corresponding feature. In the above manner, in a process of processing the multi-behavior sequences, original time information of each behavior in the multi-behavior sequences is obfuscated. As such, the multi-behavior sequence obtained through modeling is relatively different from the original behavior sequence (unprocessed multi-behavior sequence). This results in inaccurate time and type of a future behavior predicted by using the modeled multi-behavior sequence. Therefore, there is a need to provide a more accurate method for modeling a multi-behavior sequence.

SUMMARY

[0004] One or more implementations of the present specification provide methods and

apparatuses for training a behavior prediction model, and the methods can accurately model time point information of each behavior in a multi-behavior sequence of a user.

[0005] According to a first aspect, a method for training a behavior prediction model is provided, where the method includes: splitting a behavior sequence of a user to obtain multiple single-behavior sequences, each single-behavior sequence being corresponding to one behavior, and being used to record a correspondence between the behavior and a time point; performing time coding on the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences; inputting the multiple single-behavior time sequences into a behavior prediction model, so the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain a distribution situation of behaviors corresponding to the multiple single-behavior time sequences in terms of time; and training the behavior prediction model based on a first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences, the behavior prediction model being used to predict a time point and a type of at least one future behavior of the user based on the behavior sequence of the user.

[0006] In the above technical solution, in one or more implementations of the present specification, a behavior sequence of a user is split to obtain multiple single-behavior sequences; time coding is performed on all time points in the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences; the multiple single-behavior time sequences are input into a behavior prediction model, to obtain a distribution situation, predicted by the behavior prediction model, of behaviors corresponding to the multiple single-behavior time sequences in terms of time; and the behavior prediction model is trained based on a first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences. That is, the behavior prediction model pays attention to each time point of multiple single behaviors in the behavior sequence of the user, and models each time point. Therefore, a more accurate behavior prediction model can be obtained through modeling, and a more accurate time point and type of a future behavior can be predicted by using the behavior prediction model.

[0007] With reference to the first aspect, in some possible implementations, the splitting the behavior sequence of the user to obtain the multiple single-behavior sequences includes: splitting

the behavior sequence of the user according to behavior types to obtain the multiple single-behavior sequences, where the behavior sequence of the user corresponds to multiple behavior types.

5 **[0008]** In the above technical solution, because the behavior sequence of the user is corresponding to multiple behavior types, the behavior sequence of the user can be split into multiple single-behavior sequences according to the behavior types, and each single behavior in multiple single behaviors is corresponding to one behavior type. Therefore, time coding can be performed on the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences, so the multiple single-behavior time
10 sequences can be used to train the behavior prediction model.

[0009] With reference to the first aspect and the above implementation, in some possible implementations, the performing time coding on the multiple single-behavior sequences to obtain the multiple single-behavior time sequences of the multiple single-behavior sequences includes: performing time coding on the multiple single-behavior sequences by using a trigonometric
15 function, to obtain the multiple single-behavior time sequences.

[0010] In the above technical solution, time coding is performed on the multiple single-behavior sequences by using the trigonometric function, and all time features of each single-behavior sequence in the multiple single-behavior sequences in terms of time can be obtained, so as to obtain multiple single-behavior time sequences of the multiple single-behavior sequences.

20 **[0011]** With reference to the first aspect and the above implementations, in some possible implementations, that the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time includes: The behavior prediction model performs modeling of a
25 strength value of the time point in a predetermined time period for each single-behavior time sequence in the multiple single-behavior time sequences, to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0012] It should be understood that the above technical solution is to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms
30 of time, and is actually to obtain a probability density distribution function of the behaviors corresponding to the multiple single-behavior time sequences in a time period.

[0013] With reference to the first aspect and the above implementations, in some possible implementations, the behavior prediction model includes a behavior time sub-model, and the inputting the multiple single-behavior time sequences into the behavior prediction model, so the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time includes: inputting the multiple single-behavior time sequences into the behavior time sub-model in the behavior prediction model, and performing, by the behavior time sub-model, modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0014] In the above technical solution, in a case in which the behavior prediction model includes the behavior time sub-model, the behavior time sub-model is specifically trained by using the first loss value between the multiple single-behavior time sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time. The distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time is obtained by the behavior time sub-model by modeling the time point for the multiple single-behavior time sequences.

[0015] With reference to the first aspect and the above implementations, in some possible implementations, the training the behavior prediction model based on the first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences includes: training the behavior time sub-model in the behavior prediction model based on the first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences.

[0016] In the above technical solution, the behavior prediction model is trained based on the first loss value. As such, the obtained behavior prediction model can better predict time and a type of a future behavior based on a behavior sequence to be predicted.

[0017] With reference to the first aspect and the above implementations, in some possible implementations, the behavior prediction model further includes a behavior relationship sub-model and a prediction sub-model, and the method further includes: inputting the multiple single-

behavior sequences into the behavior relationship sub-model to obtain causality between every two single-behavior sequences in the multiple single-behavior sequences, where the causality is used to represent a relationship between occurrence of a behavior in one single-behavior sequence and occurrence of a behavior in another single-behavior sequence in every two single-behavior sequences; inputting, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, and performing, by the prediction sub-model, sequence prediction based on the causality and the distribution situation to obtain a predicted behavior sequence; and training the behavior relationship sub-model and the prediction sub-model based on the first loss value and a second loss value between the predicted behavior sequence and the behavior sequence of the user.

[0018] In the above technical solution, when the behavior prediction model further includes the behavior relationship sub-model and the prediction sub-model, the multiple single-behavior sequences can be input into the behavior relationship sub-model, to obtain the causality between every two single-behavior sequences in the multiple single-behavior sequences; and the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time are input into the prediction sub-model, so as to obtain the predicted behavior sequence by using the prediction sub-model. Therefore, the behavior relationship sub-model and the prediction sub-model are trained based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user. That is, the behavior prediction model not only pays attention to the time point information of the behavior sequence of the user, but also pays attention to causality between multiple behaviors in the behavior sequence of the user, so the behavior sequence of the user can be more accurately modeled.

[0019] With reference to the first aspect and the above implementations, in some possible implementations, the inputting the multiple single-behavior sequences into the behavior relationship sub-model to obtain the causality between every two single-behavior sequences in the multiple single-behavior sequences includes: obtaining the causality between every two single-behavior sequences in the multiple single-behavior sequences according to a sum of covariance values of every two single-behavior sequences in the multiple single-behavior sequences at all

time points.

[0020] It should be understood that the above obtaining the causality between every two single-behavior sequences in the multiple single-behavior sequences is determining Granger causality between every two single-behavior sequences.

5 **[0021]** With reference to the first aspect and the above implementations, in some possible implementations, the training the behavior relationship sub-model and the prediction sub-model based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user includes: determining a third loss value between a prediction label and a real label of the behavior sequence of the user, where the real label is used to indicate
10 real information of the user, and the prediction label is used to indicate information of the user predicted by the prediction sub-model; obtaining a fourth loss value by averaging the second loss value and the third loss value; and training the behavior relationship sub-model and the prediction sub-model based on the first loss value and the fourth loss value.

[0022] In the above technical solution, in comparison with training of the behavior relationship
15 sub-model and the prediction sub-model based only on the first loss value and the second loss value, the above technical solution further introduces the third loss value between the prediction label and the real label. The average value between the third loss value and the second loss value is determined as the fourth loss value, and the behavior relationship sub-model and the prediction sub-model are trained based on the first loss value and the fourth loss value. In this solution, the
20 third loss value is introduced, which is equivalent to introducing more information to train the behavior relationship sub-model and the prediction sub-model. As such, the obtained behavior relationship sub-model and prediction sub-model are more accurate.

[0023] With reference to the first aspect and the above implementations, in some possible implementations, the method further includes: before the determining the third loss value between
25 the prediction label and the real label of the behavior sequence of the user, inputting, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, to obtain the prediction label corresponding to the behavior sequence of the user.

30 **[0024]** With reference to the first aspect and the above implementations, in some possible implementations, the training the behavior relationship sub-model and the prediction sub-model

based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user includes: obtaining a fifth loss value according to the first loss value and the second loss value; and training the behavior relationship sub-model based on the fifth loss value, and training the prediction sub-model based on the second loss value.

5 **[0025]** With reference to the first aspect and the above implementations, in some possible implementations, the obtaining the fifth loss value according to the first loss value and the second loss value includes: determining a first value according to a matrix formed by a second-order partial derivation of each weight in the relationship sub-model and a vector formed by each weight in the relationship sub-model; and obtaining the fifth loss value according to the first value, the
10 first loss value, and the second loss value.

[0026] It should be understood that the above process of determining the first value is implemented by using a Pareto optimality solution principle.

[0027] With reference to the first aspect and the above implementations, in some possible implementations, the method further includes: inputting a behavior sequence to be predicted into
15 the behavior prediction model to obtain a time point and a type corresponding to at least one future behavior of the behavior sequence to be predicted as predicted by the behavior prediction model.

[0028] In summary, one or more implementations of the present specification provide a method for training a behavior prediction model. A behavior sequence of a user is split to obtain multiple single-behavior sequences; time coding is performed on all time points in the multiple
20 single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences; the multiple single-behavior time sequences are input into a behavior prediction model, to obtain a distribution situation, predicted by the behavior prediction model, of behaviors corresponding to the multiple single-behavior time sequences in terms of time; and the behavior prediction model is trained based on a first loss value between the distribution situation
25 of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences. That is, the behavior prediction model pays attention to each piece of time point information of multiple single behaviors in the behavior sequence of the user, and models each time point. However, to avoid the original time information of each behavior in the multi-behavior sequence from being obfuscated in a related technology, a
30 more accurate behavior prediction model can be obtained through modeling, and the behavior prediction model can be used to predict a more accurate time point and type of a future behavior.

[0029] In addition, when the behavior prediction model includes the behavior time sub-model, the behavior relationship sub-model, and the prediction sub-model, the multiple single-behavior sequences can be input into the behavior relationship sub-model, to obtain the causality between every two single-behavior sequences in the multiple single-behavior sequences; and the causality
5 between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time are input into the prediction sub-model, so as to obtain the predicted user sequence by using the prediction sub-model. Therefore, the behavior relationship sub-model and the prediction sub-model are trained based on the first loss value and the second loss value
10 between the predicted behavior sequence and the behavior sequence of the user. That is, the behavior prediction model not only pays attention to the time point information of the behavior sequence of the user, but also pays attention to causality between multiple behaviors in the behavior sequence of the user, so the behavior sequence of the user can be more accurately modeled.

[0030] Finally, a time point and a type corresponding to at least one future behavior of the behavior sequence to be predicted can be predicted by using the obtained behavior prediction model.

[0031] According to a second aspect, an apparatus for training a behavior prediction model is provided, where the apparatus includes: a determining module, configured to split a behavior
20 sequence of a user to obtain multiple single-behavior sequences, each single-behavior sequence being corresponding to one behavior, and being used to record a correspondence between the behavior and a time point; the determining module being further configured to perform time coding on the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences; the determining module being further configured to input
25 the multiple single-behavior time sequences into a behavior prediction model, so the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain a distribution situation of behaviors corresponding to the multiple single-behavior time sequences in terms of time; and a training module, configured to train the behavior prediction model based on a first loss value between the
30 distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences, the behavior

prediction model being used to predict a time point and a type of at least one future behavior of the user based on the behavior sequence of the user.

[0032] With reference to the second aspect, in some possible implementations, the determining module is configured to split the behavior sequence of the user according to behavior types to obtain the multiple single-behavior sequences, where the behavior sequence of the user corresponds to multiple behavior types.

[0033] With reference to the second aspect and the above implementation, in some possible implementations, the determining module is further configured to perform time coding on the multiple single-behavior sequences by using a trigonometric function, to obtain the multiple single-behavior time sequences.

[0034] With reference to the second aspect and the above implementations, in some possible implementations, the determining module is further configured to: perform, by using the behavior prediction model, modeling of a strength value of the time point in a predetermined time period for each single-behavior time sequence in the multiple single-behavior time sequences, to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0035] It should be understood that the above solution is to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, and is actually to obtain a probability density distribution function of the behaviors corresponding to the multiple single-behavior time sequences in a time period.

[0036] With reference to the second aspect and the above implementations, in some possible implementations, the behavior prediction model includes a behavior time sub-model, and the determining module is specifically further configured to: input the multiple single-behavior time sequences into the behavior time sub-model in the behavior prediction model, and perform, by the behavior time sub-model, modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0037] With reference to the second aspect and the above implementations, in some possible implementations, the training module is configured to train the behavior time sub-model in the behavior prediction model based on the first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the

multiple single-behavior time sequences.

[0038] With reference to the second aspect and the above implementations, in some possible implementations, the behavior prediction model further includes a behavior relationship sub-model and a prediction sub-model, and the determining module is further configured to: input the multiple single-behavior sequences into the behavior relationship sub-model to obtain causality between every two single-behavior sequences in the multiple single-behavior sequences, where the causality is used to represent a relationship between occurrence of a behavior in one single-behavior sequence and occurrence of a behavior in another single-behavior sequence in every two single-behavior sequences; and input, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, and perform, by the prediction sub-model, sequence prediction based on the causality and the distribution situation to obtain a predicted behavior sequence; and the training module is further configured to: train the behavior relationship sub-model and the prediction sub-model based on the first loss value and a second loss value between the predicted behavior sequence and the behavior sequence of the user.

[0039] With reference to the second aspect and the above implementations, in some possible implementations, the determining module is further configured to obtain the causality between every two single-behavior sequences in the multiple single-behavior sequences according to a sum of covariance values of every two single-behavior sequences in the multiple single-behavior sequences at all time points.

[0040] With reference to the second aspect and the above implementations, in some possible implementations, the training module is further configured to: determine a third loss value between a prediction label and a real label of the behavior sequence of the user, where the real label is used to indicate real information of the user, and the prediction label is used to indicate information of the user predicted by the prediction sub-model; obtain a fourth loss value by averaging the second loss value and the third loss value; and train the behavior relationship sub-model and the prediction sub-model based on the first loss value and the fourth loss value.

[0041] With reference to the second aspect and the above implementations, in some possible implementations, the determining module is further configured to: input, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior

sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, to obtain the prediction label corresponding to the behavior sequence of the user.

[0042] With reference to the second aspect and the above implementations, in some possible implementations, the training module is further configured to: obtain a fifth loss value according to the first loss value and the second loss value; and train the behavior relationship sub-model based on the fifth loss value, and train the prediction sub-model based on the second loss value.

[0043] With reference to the second aspect and the above implementations, in some possible implementations, the determining module is further configured to: determine a first value according to a matrix formed by a second-order partial derivation of each weight in the relationship sub-model and a vector formed by each weight in the relationship sub-model; and obtain the fifth loss value according to the first value, the first loss value, and the second loss value.

[0044] It should be understood that the above process of determining the first value is implemented by using a Pareto optimality solution principle.

[0045] With reference to the second aspect and the above implementations, in some possible implementations, the determining module is further configured to: input a behavior sequence to be predicted into the behavior prediction model to obtain a time point and a type corresponding to at least one future behavior of the behavior sequence to be predicted as predicted by the behavior prediction model.

[0046] According to a third aspect, an electronic device is provided, including a memory, a processor, and a computer program stored in the memory and running on the processor, where when the processor executes the computer program, the electronic device is enabled to perform the method in any one of the first aspect or the possible implementations of the first aspect.

[0047] According to a fourth aspect, a computer-readable storage medium is provided, where the computer-readable storage medium stores instructions, and when the instructions run on a computer or a processor, the computer or the processor is enabled to perform the method in any one of the first aspect or the possible implementations of the first aspect.

[0048] According to a fifth aspect, a computer program product including instructions is provided. When the computer program product runs on a computer or a processor, the computer or the processor is enabled to perform the method in any one of the first aspect or the possible implementations of the first aspect.

BRIEF DESCRIPTION OF DRAWINGS

[0049] FIG. 1 is a schematic diagram of performing time aggregation on a multi-behavior sequence according to one or more implementations of the present specification.

[0050] FIG. 2 is a schematic flowchart illustrating a method for training a behavior prediction model according to one or more implementations of the present specification.

[0051] FIG. 3 is a schematic diagram of splitting a behavior sequence of a user according to one or more implementations of the present specification.

[0052] FIG. 4 is a schematic structural diagram illustrating a behavior prediction model according to one or more implementations of the present specification.

[0053] FIG. 5 is a schematic structural diagram illustrating an apparatus for training a behavior prediction model according to one or more implementations of the present specification.

[0054] FIG. 6 is a schematic structural diagram illustrating an electronic device, according to one or more implementations of the present specification.

DESCRIPTION OF EMBODIMENTS

[0055] The following clearly and thoroughly describes the technical solutions in one or more implementations of the present specification with reference to the accompanying drawings. In the description of one or more implementations of the present specification, "multiple" refers to two or more than two. The terms "first" and "second" are merely used for description, and shall not be understood as an indication or implication of relative importance or implicit indication of the number of indicated technical features. Therefore, a feature limited by "first" or "second" can explicitly or implicitly include one or more such features. "and/or" in the present specification is merely an association relationship between associated objects, and represents that three relationships can exist. For example, A and/or B can represent the following three cases: Only A exists, both A and B exist, and only B exists.

[0056] FIG. 1 is a schematic diagram of performing time aggregation on a multi-behavior sequence according to one or more implementations of the present specification.

[0057] For example, in a multi-behavior sequence shown in FIG. 1, the multi-behavior sequence is corresponding to multiple behavior types, and each behavior type is corresponding to a time point. As shown in FIG. 1, at a time point of the 0th minute, it is a pentagonal type of

behavior; at a time point of the 1st minute, it is a circular type of behavior; at a time point of the 2.2nd minute, it is a triangular type of behavior; at a time point of the 3rd minute, it is a triangular type of behavior; at a time point of the 3.5th minute, it is a pentagonal type of behavior; and by analogy, at the last time point of the 27th minute, it is a rectangular type of behavior. The pentagonal type is an online-purchase type, the circular type is a transfer type, the triangular type is a fund subscription type, and the rectangular type is a chat type.

[0058] In a related technology, for the multi-behavior sequence shown in FIG. 1, the multi-behavior sequence is equally divided according to a predetermined time interval, to obtain behavior sequences in multiple even time periods. Time aggregation is performed on behavior sequences in each time period of the behavior sequences in the multiple even time periods, and a quantity of times of occurrence of each behavior in each time period is obtained. Assume that the predetermined time interval is 4 minutes, and the following is obtained after the multi-behavior sequence is equally divided and time aggregation: 0–4 minutes: 2 online purchase behaviors, 1 transfer behavior, 2 fund subscription behaviors, and 0 chat behaviors; 4–8 minutes: 0 online purchases, 1 transfer behavior, 1 fund subscription behavior, and 2 chat behaviors; 8–12 minutes: 2 online purchase behaviors, 1 transfer behavior, 0 fund subscription behaviors, and 1 chat behavior; 12–16 minutes: 1 online purchase behavior, 1 transfer behavior, 1 fund subscription behavior, and 1 chat behavior; 16–20 minutes: 1 online purchase behavior, 2 transfer behaviors, 0 fund subscription behaviors, and 0 chat behaviors; 20–24 minutes: 0 online purchase behaviors, 0 transfer behaviors, 2 fund subscription behaviors, and 1 chat behavior; and 24–28 minutes: 1 online purchase behavior, 0 transfer behaviors, 2 fund subscription behaviors, and 2 chat behaviors.

[0059] According to the rule of the multi-behavior sequence collected every 4 minutes, the prediction yields a time period of occurrence of a specific type of behavior in the future, rather than a time point. Therefore, the prediction manner is inaccurate in predicting time information of occurrence of a future behavior.

[0060] FIG. 2 is a schematic flowchart illustrating a method for training a behavior prediction model according to one or more implementations of the present specification.

[0061] It should be understood that the method for training a behavior prediction model provided in one or more implementations of the present specification can be applied to any electronic device, and the electronic device can be a computer terminal (a display device or a mobile phone terminal having a data processing function), a server, etc.

[0062] It should be further understood that, in one or more implementations of the present specification, a process of determining the behavior prediction model is implemented based on a recurrent neural network (RNN) in deep learning, which can be, for example, a long short term memory (LSTM) unit, a gated recurrent unit (GRU), etc.

5 **[0063]** For example, as shown in FIG. 2, the method 200 includes step 202 to step 208.

[0064] In step 202, the electronic device splits a behavior sequence of a user to obtain multiple single-behavior sequences, each single-behavior sequence being corresponding to one behavior, and being used to record a correspondence between the behavior and a time point.

[0065] It should be understood that the behavior sequence of the user in the above solution
10 refers to the multi-behavior sequence shown in FIG. 1, and same as the multi-behavior sequence, the behavior sequence of the user corresponds to multiple types of behavior.

[0066] It should be further understood that each behavior of a type in the behavior sequence of the user has a corresponding time point, that is, specific occurrence time of a specific behavior of a type.

15 **[0067]** In a possible implementation, step 202 includes: The electronic device splits the behavior sequence of the user according to behavior types to obtain the multiple single-behavior sequences, where the behavior sequence of the user corresponds to multiple behavior types.

[0068] In the above technical solution, because the behavior sequence of the user is corresponding to multiple behavior types, the behavior sequence of the user can be split into
20 multiple single-behavior sequences according to the behavior types, and each single behavior in multiple single behaviors is corresponding to one behavior type. Therefore, time coding can be performed on the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences, so the multiple single-behavior time sequences can be used to train the behavior prediction model.

25 **[0069]** FIG. 3 is a schematic diagram of splitting a behavior sequence of a user according to one or more implementations of the present specification.

[0070] For example, FIG. 3 specifically shows multiple single-behavior sequences obtained by splitting the multi-behavior sequence shown in FIG. 1 (e.g., the behavior sequence of the user with a sequence number ①). The multiple behavior types corresponding to the behavior sequence
30 of the user are a pentagonal type, namely, an online purchase type, a circular type, namely, a transfer type, a triangular type, namely, a fund subscription type, and a rectangular type, namely,

a chat type. According to the behavior types, the behavior sequence of the user is split into multiple single-behavior sequences, that is, a sequence with a sequence number ②, a sequence with a sequence number ③, a sequence with a sequence number ④, and a sequence with a sequence number ⑤, that is, the behavior sequence of the user is split into a behavior sequence of the online purchase type, a behavior sequence of the transfer type, a behavior sequence of the fund subscription type, and a behavior sequence of the chat type.

[0071] In step 204, the electronic device performs time coding on the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences.

[0072] In a possible implementation, step 204 includes: The electronic device performs time coding on the multiple single-behavior sequences by using a trigonometric function, to obtain the multiple single-behavior time sequences.

[0073] In the above technical solution, time coding is performed on the multiple single-behavior sequences by using the trigonometric function, and all time features of each single-behavior sequence in the multiple single-behavior sequences in terms of time can be obtained, so as to obtain multiple single-behavior time sequences of the multiple single-behavior sequences.

[0074] For example, the electronic device performs, by using the following equation (1) of the trigonometric function, time coding on each single-behavior sequence in the multiple single-behavior sequences, to obtain the multiple single-behavior time sequences.

$$\Phi_T(t - \tau) = \varepsilon(t, \tau) = \frac{1}{\sqrt{d}} [\cos(\omega_1(t - \tau) + b_1), \dots, \cos(\omega_d(t - \tau) + b_d)] \quad (1)$$

[0075] Φ_T indicates a mapping of a behavior sequence within a time period; $b_1, \dots, b_d$ are phase difference parameters; $\omega_1, \dots, \omega_d$ are frequency parameters; d is a dimension obtained by mapping a behavior sequence by using a trigonometric function; and $t - \tau$ indicates a time period.

[0076] 206. The electronic device inputs the multiple single-behavior time sequences into a behavior prediction model, so the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain a distribution situation of behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0077] In a possible implementation, step 206 includes: The behavior prediction model performs modeling of a strength value of the time point in a predetermined time period for each single-behavior time sequence in the multiple single-behavior time sequences, to obtain the

distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0078] It should be understood that the above technical solution is to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, and is actually to obtain a probability density distribution function of the behaviors corresponding to the multiple single-behavior time sequences in a time period.

[0079] It should be understood that, the behavior prediction model performs modeling of the strength value of the time point in the predetermined time period for each single-behavior time sequence in the multiple single-behavior time sequences, to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time. That is, a probability density distribution function, as represented by the following equation (2), of the behavior corresponding to each single-behavior time sequence in the multiple single-behavior time sequences in a time period is obtained.

$$p^*(\tau_i) = p(\tau_i | H_{t_i}) = \lambda^*(t_{i-1} + \tau_i) \exp\left(-\int_0^{\tau_i} \lambda^*(t_{i-1} + s) ds\right) \quad (2)$$

[0080] τ_i indicates a time period; $p^*(\tau_i)$ is a probability density distribution function of a behavior corresponding to a single-behavior time sequence in a time period; $p^*(\tau_i) = p(\tau_i | H_{t_i})$ indicates that a probability density distribution function of a behavior corresponding to a single-behavior time sequence in a time period is related to a historical behavior of the single-behavior time sequence; $\lambda^*(t_{i-1} + \tau_i)$ is a strength function (strength value) of occurrence of a behavior in a time period from t_{i-1} to τ_i , that is, a probability of occurrence of the behavior; and s indicates a time point in the time period from t_{i-1} to τ_i .

[0081] In a possible implementation, in a case in which the behavior prediction model includes a behavior time sub-model, step 206 includes: The electronic device inputs the multiple single-behavior time sequences into the behavior time sub-model in the behavior prediction model, so the behavior time sub-model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0082] In the above technical solution, a more specific implementation of step 206 is specifically as follows: In a case in which the behavior prediction model includes the behavior time sub-model, the multiple single-behavior time sequences can be input into the behavior time

sub-model, and the behavior time sub-model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences.

[0083] 208. The electronic device trains the behavior prediction model based on a first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences, the behavior prediction model being used to predict a time point and a type of at least one future behavior of the user based on the behavior sequence of the user.

[0084] In a possible implementation, in a case in which the behavior prediction model includes a behavior time sub-model, step 208 includes: The electronic device trains the behavior time sub-model in the behavior prediction model based on the first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences.

[0085] In the above technical solution, in a case in which the behavior prediction model includes the behavior time sub-model, the behavior time sub-model is specifically trained by using the first loss value between the multiple single-behavior time sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time. The distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time is obtained by the behavior time sub-model by modeling the time point for the multiple single-behavior time sequences.

[0086] In a possible implementation, in a case in which the behavior prediction model further includes a behavior relationship sub-model and a prediction sub-model, the method further includes: The electronic device inputs the multiple single-behavior sequences into the behavior relationship sub-model to obtain causality between every two single-behavior sequences in the multiple single-behavior sequences, where the causality is used to represent a relationship between occurrence of a behavior in one single-behavior sequence and occurrence of a behavior in another single-behavior sequence in every two single-behavior sequences; the electronic device inputs, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, so the prediction sub-model performs sequence prediction based on the causality and the distribution situation to obtain a predicted behavior sequence; and the electronic device trains the behavior relationship sub-model

and the prediction sub-model based on the first loss value and a second loss value between the predicted behavior sequence and the behavior sequence of the user.

[0087] In the above technical solution, when the behavior prediction model further includes the behavior relationship sub-model and the prediction sub-model, the multiple single-behavior sequences can be input into the behavior relationship sub-model, to obtain the causality between every two single-behavior sequences in the multiple single-behavior sequences; and the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time are input into the prediction sub-model, so as to obtain the predicted user sequence by using the prediction sub-model. Therefore, the behavior relationship sub-model and the prediction sub-model are trained based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user. That is, the behavior prediction model not only pays attention to the time point information of the behavior sequence of the user, but also pays attention to causality between multiple behaviors in the behavior sequence of the user, so the behavior sequence of the user can be more accurately modeled.

[0088] In a possible implementation, that the electronic device inputs the multiple single-behavior sequences into the behavior relationship sub-model to obtain the causality between every two single-behavior sequences in the multiple single-behavior sequences includes: The electronic device obtains the causality between every two single-behavior sequences in the multiple single-behavior sequences according to a sum of covariance values of every two single-behavior sequences in the multiple single-behavior sequences at all time points.

[0089] It should be understood that the above obtaining the causality between every two single-behavior sequences in the multiple single-behavior sequences is specifically determining Granger causality between every two single-behavior sequences.

[0090] It should be further understood that the electronic device obtains, according to the following equation (3) with the Granger causality, the causality between every two single-behavior sequences in the multiple single-behavior sequences.

$$f_k(X_i) - f_k(\bar{X}_i) = \sum_{j=1}^i A_j (f_k, X_i, \bar{X}_i) \quad (3)$$

[0091] X_i and $\bar{X}_i$ are respectively any two different single-behavior sequences in the multiple single-behavior sequences; i is an i th behavior in a single-behavior sequence; any two

single-behavior sequences in the multiple single-behavior sequences can form a group of sequences in k groups; $f_k(X_i) - f_k(\tilde{X}_i)$ indicates Granger causality between two single-behavior sequences in a k th group of sequences; and $A_j(f_k, X_i, \tilde{X}_i)$ indicates a covariance of an i th behavior between two single-behavior sequences in the k th group of sequences.

5 **[0092]** In a possible implementation, that the electronic device trains the behavior relationship sub-model and the prediction sub-model based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user includes: The electronic device determines a third loss value between a prediction label and a real label of the behavior sequence of the user, where the real label is used to indicate real information of the user,
10 and the prediction label is used to indicate information of the user predicted by the prediction sub-model; the electronic device obtains a fourth loss value by averaging the second loss value and the third loss value; and the electronic device trains the behavior relationship sub-model and the prediction sub-model based on the first loss value and the fourth loss value.

[0093] In the above technical solution, in comparison with training of the behavior relationship
15 sub-model and the prediction sub-model based only on the first loss value and the second loss value, the above technical solution further introduces the third loss value between the prediction label and the real label. The average value between the third loss value and the second loss value is determined as the fourth loss value, and the behavior relationship sub-model and the prediction sub-model are trained based on the first loss value and the fourth loss value. In this solution, the
20 third loss value is introduced, which is equivalent to introducing more information to train the behavior relationship sub-model and the prediction sub-model. As such, the obtained behavior relationship sub-model and prediction sub-model are more accurate.

[0094] In a possible implementation, before the electronic device determines the third loss value between the prediction label and the real label of the behavior sequence of the user, the
25 method further includes: The electronic device inputs, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, to obtain the prediction label corresponding to the behavior sequence of the user.

30 **[0095]** Optionally, the real label of the behavior sequence of the user can be identity information and/or occupation information of the user.

[0096] The user's consent needs to be obtained for obtaining the identity information and/or the occupation information of the user.

[0097] In a possible implementation, that the electronic device trains the behavior relationship sub-model and the prediction sub-model based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user includes: The electronic device obtains a fifth loss value according to the first loss value and the second loss value; and the electronic device trains the behavior relationship sub-model based on the fifth loss value, and trains the prediction sub-model based on the second loss value.

[0098] In a possible implementation, that the electronic device obtains the fifth loss value according to the first loss value and the second loss value includes: The electronic device determines a first value according to a matrix formed by a second-order partial derivation of each weight in the relationship sub-model and a vector formed by each weight in the relationship sub-model; and the electronic device obtains the fifth loss value according to the first value, the first loss value, and the second loss value.

[0099] It should be understood that the above process of determining the first value is implemented by using a Pareto optimality solution principle.

[0100] Specifically, the electronic device obtains the first value according to the following Pareto optimality equation (4).

$$Hv = \nabla f \beta \in \text{cplspan}\{\nabla f_i\} \quad (4)$$

[0101] H is a matrix formed by a second-order partial derivation of each weight in the behavior relationship sub-model, and is specifically a Hessian matrix; v is the first value; f is a vector value of the behavior relationship sub-model; ∇f_i is a target gradient of each weight; and β is a vector value formed by each weight in the behavior relationship sub-model.

[0102] In a possible implementation, the method further includes: The electronic device inputs a behavior sequence to be predicted into the behavior prediction model to obtain a time point and a type corresponding to at least one future behavior of the behavior sequence to be predicted as predicted by the behavior prediction model.

[0103] The above technical solution is specifically a process of using the behavior prediction model obtained through training. A time point and a type corresponding to at least one future behavior of any behavior sequence can be obtained by predicting the behavior sequence by using the trained behavior prediction model.

[0104] In a possible implementation, the electronic device splits a user sequence to be predicted according to behavior types to obtain multiple single-behavior user sequences. The electronic device performs time coding on the multiple single-behavior user sequences to obtain multiple single-behavior user time sequences, where the behavior sequence to be predicted includes multiple single-behavior user sequences and multiple single-behavior user time sequences.

[0105] Optionally, that the electronic device performs time coding on the multiple single-behavior user sequences to obtain the multiple single-behavior user time sequences includes: The electronic device performs time coding on the multiple single-behavior user sequences by using a trigonometric function, to obtain the multiple single-behavior user time sequences.

[0106] In a possible implementation, the electronic device inputs the multiple single-behavior user time sequences into a behavior time sub-model to obtain a distribution situation of behaviors corresponding to the multiple single-behavior user time sequences in terms of time. The electronic device inputs the multiple single-behavior user sequences into a behavior relationship sub-model, to obtain causality between every two single-behavior user sequences in the multiple single-behavior user sequences. The electronic device inputs the distribution situation and the causality into a prediction sub-model to obtain a time point and a type corresponding to at least one future behavior of the behavior sequence to be predicted.

[0107] FIG. 4 is a schematic structural diagram illustrating a behavior prediction model according to one or more implementations of the present specification.

[0108] As shown in FIG. 4, a process of obtaining a behavior prediction model is specifically described. As shown in FIG. 4, specifically, a behavior sequence of a user is split to obtain multiple single-behavior sequences. Time coding is performed on the multiple single-behavior sequences to obtain multiple single-behavior time sequences. A behavior time sub-model performs modeling of time points in the multiple single-behavior time sequences to obtain a distribution situation of behaviors corresponding to the multiple single-behavior time sequences in terms of time. The multiple single-behavior sequences are input into a behavior relationship sub-model, and the behavior relationship sub-model obtains causality between every two single-behavior sequences in the multiple single-behavior sequences. A prediction sub-model can predict a predicted behavior sequence and a prediction label according to the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the causality between every two single-behavior sequences in the multiple single-behavior time sequences.

[0109] A first loss value between the multiple single-behavior time sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time is determined; a second loss value between the behavior sequence of the user and the predicted behavior sequence is determined; and a third loss value between a real label and the prediction label is determined.

[0110] The behavior time sub-model, the behavior relationship sub-model, and the prediction sub-model can be trained based on the first loss value and the second loss value. Or a fourth loss value can be obtained based on an average value of the second loss value and the third loss value, and the behavior time sub-model, the behavior relationship sub-model, and the prediction sub-model are trained based on the first loss value and the fourth loss value.

[0111] FIG. 5 is a schematic structural diagram illustrating an apparatus for training a behavior prediction model according to one or more implementations of the present specification.

[0112] For example, as shown in FIG. 5, the apparatus 500 includes: a determining module 501, configured to split a behavior sequence of a user to obtain multiple single-behavior sequences, each single-behavior sequence being corresponding to one behavior, and being used to record a correspondence between the behavior and a time point; the determining module 501 being further configured to perform time coding on the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences; the determining module 501 being further configured to input the multiple single-behavior time sequences into a behavior prediction model, so the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain a distribution situation of behaviors corresponding to the multiple single-behavior time sequences in terms of time; and a training module 502, configured to train the behavior prediction model based on a first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences, the behavior prediction model being used to predict a time point and a type of at least one future behavior of the user based on the behavior sequence of the user.

[0113] Optionally, the determining module 501 is configured to split the behavior sequence of the user according to behavior types to obtain the multiple single-behavior sequences, where the behavior sequence of the user corresponds to multiple behavior types.

[0114] Optionally, the determining module 501 is further configured to perform time coding

on the multiple single-behavior sequences by using a trigonometric function, to obtain the multiple single-behavior time sequences.

[0115] Optionally, the determining module 501 is further configured to: perform, by using the behavior prediction model, modeling of a strength value of the time point in a predetermined time period for each single-behavior time sequence in the multiple single-behavior time sequences, to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0116] It should be understood that the above solution is to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, and is actually to obtain a probability density distribution function of the behaviors corresponding to the multiple single-behavior time sequences in a time period.

[0117] Optionally, the behavior prediction model includes a behavior time sub-model, and the determining module 501 is further configured to: input the multiple single-behavior time sequences into the behavior time sub-model in the behavior prediction model, and perform, by the behavior time sub-model, modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

[0118] Optionally, the training module 502 is configured to train the behavior time sub-model in the behavior prediction model based on the first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences.

[0119] Optionally, the behavior prediction model further includes a behavior relationship sub-model and a prediction sub-model, and the determining module 501 is further configured to: input the multiple single-behavior sequences into the behavior relationship sub-model to obtain causality between every two single-behavior sequences in the multiple single-behavior sequences, where the causality is used to represent a relationship between occurrence of a behavior in one single-behavior sequence and occurrence of a behavior in another single-behavior sequence in every two single-behavior sequences; and input, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, and perform, by the prediction sub-model, sequence prediction based on the causality and

the distribution situation to obtain a predicted behavior sequence; and the training module 502 is further configured to: train the behavior relationship sub-model and the prediction sub-model based on the first loss value and a second loss value between the predicted behavior sequence and the behavior sequence of the user.

5 **[0120]** Optionally, the determining module 501 is further configured to obtain the causality between every two single-behavior sequences in the multiple single-behavior according to a sum of covariance values of every two single-behavior sequences in the multiple single-behavior sequences at all time points.

10 **[0121]** Optionally, the training module 502 is further configured to: determine a third loss value between a prediction label and a real label of the behavior sequence of the user, where the real label is used to indicate real information of the user, and the prediction label is used to indicate information of the user predicted by the prediction sub-model; obtain a fourth loss value by averaging the second loss value and the third loss value; and train the behavior relationship sub-model and the prediction sub-model based on the first loss value and the fourth loss value.

15 **[0122]** Optionally, the determining module 501 is further configured to: input, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, to obtain the prediction label corresponding to the behavior sequence of the user.

20 **[0123]** Optionally, the training module 502 is further configured to: obtain a fifth loss value according to the first loss value and the second loss value; and train the behavior relationship sub-model based on the fifth loss value, and train the prediction sub-model based on the second loss value.

25 **[0124]** Optionally, the determining module 501 is further configured to: determine a first value according to a matrix formed by a second-order partial derivation of each weight in the relationship sub-model and a vector formed by each weight in the relationship sub-model; and obtain the fifth loss value according to the first value, the first loss value, and the second loss value.

[0125] It should be understood that the above process of determining the first value is implemented by using a Pareto optimality solution principle.

30 **[0126]** Optionally, the determining module 501 is further configured to: input a behavior sequence to be predicted into the behavior prediction model to obtain a time point and a type

corresponding to at least one future behavior of the behavior sequence to be predicted as predicted by the behavior prediction model.

[0127] FIG. 6 is a schematic structural diagram illustrating an electronic device, according to one or more implementations of the present specification.

5 **[0128]** For example, as shown in FIG. 6, the electronic device 600 includes a memory 601, a processor 602, and a computer program 603 stored in the memory 601 and running on the processor 602. When the processor 602 executes the computer program 603, the electronic device can perform any above method for training a behavior prediction model.

[0129] In one or more implementations of the present specification, function modules of the electronic device can be divided according to the above method examples. For example, the division can be corresponding to each function module, or two or more functions can be integrated into one processing module. The integrated module can be implemented in a form of hardware. It should be noted that module division in one or more implementations of the present specification is an example, and is merely logical function division. In actual implementation, there can be
10 another division manner.

[0130] In a case in which each function module is divided by using a corresponding function, the electronic device can include a determining module, a training module, etc. It should be noted that all related content of each step involved in the above method implementation can be referenced to a function description of a corresponding function module, and details are omitted herein for
20 simplicity.

[0131] An electronic device provided in one or more implementations of the present specification is configured to perform the above method for training a behavior prediction model. Therefore, the same effect as that in the above implementation method can be achieved.

[0132] In a case in which an integrated unit is used, the electronic device can include a
25 processing module and a storage module. The processing module can be configured to control and manage an action of the electronic device. The storage module can be configured to support the electronic device in executing mutual program code, data, etc.

[0133] The processing module can be a processor or a controller, and can implement or execute various example logical blocks, modules, and circuits described with reference to content disclosed
30 in one or more implementations of the present specification. The processor can alternatively be a combination that implements a computing function, for example, includes a combination of one

or more microprocessors, a combination of a digital signal processor (DSP) and a microprocessor, and the storage module can be a memory.

[0134] An electronic device provided in one or more implementations of the present specification can be specifically a chip, a component, or a module, and the electronic device can include a processor and a memory connected to each other. The memory is configured to store instructions. When the electronic device runs, the processor can invoke and execute the instructions, so the chip performs any above method for training a behavior prediction model.

[0135] One or more implementations of the present specification provide a computer-readable storage medium. The computer-readable storage medium stores instructions. When the instructions run on a computer or a processor, the computer or the processor performs any above method for training a behavior prediction model.

[0136] One or more implementations of the present specification further provide a computer program product including instructions. When the computer program product runs on a computer or a processor, the computer or the processor performs the above related steps, so as to implement any above method for training a behavior prediction model.

[0137] The electronic device, the computer-readable storage medium, the computer program product including instructions, or the chip that are provided in one or more implementations of the present specification are all used to perform the above corresponding method. Therefore, for beneficial effects that can be achieved thereby, references can be made to beneficial effects in the above corresponding method. Details are omitted herein for simplicity again.

[0138] It can be learned by a person skilled in the art from the above descriptions of the implementations that for convenience and brevity of description, only division into the above functional modules is used as an example for description. In practice, the above functions can be allocated to different functional modules for implementation according to a requirement, that is, an internal structure of an apparatus is divided into different functional modules, to complete all or some of the functions described above.

[0139] In one or more implementations of the present specification, it should be understood that the disclosed apparatus and method can be implemented in another manner. For example, the described apparatus implementation is merely an example. For example, the module or division into the units is merely logical function division and can be other division in actual implementation. For example, a plurality of units or components can be combined or integrated into another

apparatus, or some features can be ignored or not performed. In addition, the displayed or discussed mutual couplings or direct couplings or communication connections can be implemented by using some interfaces. The indirect couplings or communication connections between the apparatuses or units can be implemented in electronic, mechanical, or other forms.

5 **[0140]** Specific implementations of the present specification are described above. Other implementations fall within the scope of the appended claims. In some cases, actions or steps described in the claims can be performed in an order different from that in the implementations and the desired results can still be achieved. In addition, the process depicted in the accompanying drawings does not necessarily need a particular execution order to achieve the desired results. In
10 some implementations, multi-tasking and concurrent processing are feasible or may be advantageous.

[0141] The above content is only one or more specific implementations of the present specification. However, the protection scope of the one or more implementations of the present specification is not limited thereto. Any change or replacement readily figured out by a person
15 skilled in the art within the technical scope disclosed in the one or more implementations of the present specification shall fall within the protection scope of the one or more implementations of the present specification. Therefore, the protection scope of one or more implementations of the present specification shall be subject to the protection scope of the claims.

CLAIMS

What is claimed is:

1. A method for training a behavior prediction model, comprising:

splitting a behavior sequence of a user to obtain multiple single-behavior sequences, each
5 single-behavior sequence being corresponding to one behavior, and being used to record a
correspondence between the behavior and a time point;

performing time coding on the multiple single-behavior sequences to obtain multiple
single-behavior time sequences of the multiple single-behavior sequences;

inputting the multiple single-behavior time sequences into a behavior prediction model, so
10 the behavior prediction model performs modeling of the time point for each single-behavior time
sequence in the multiple single-behavior time sequences to obtain a distribution situation of
behaviors corresponding to the multiple single-behavior time sequences in terms of time; and

training the behavior prediction model based on a first loss value between the distribution
situation of the behaviors corresponding to the multiple single-behavior time sequences in terms
15 of time and the multiple single-behavior time sequences, the behavior prediction model being used
to predict a time point and a type of at least one future behavior of the user based on the behavior
sequence of the user.

2. The method according to claim 1, wherein the splitting the behavior sequence of the user
20 to obtain the multiple single-behavior sequences includes:

splitting the behavior sequence of the user according to behavior types to obtain the
multiple single-behavior sequences, wherein the behavior sequence of the user corresponds to
multiple behavior types.

3. The method according to claim 1, wherein the performing time coding on the multiple
25 single-behavior sequences to obtain the multiple single-behavior time sequences of the multiple
single-behavior sequences includes:

performing time coding on the multiple single-behavior sequences by using a trigonometric
function, to obtain the multiple single-behavior time sequences.

4. The method according to claim 1, wherein that the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time includes:

the behavior prediction model performs modeling of a strength value of the time point in a predetermined time period for each single-behavior time sequence in the multiple single-behavior time sequences, to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

5. The method according to any one of claims 1 to 4, wherein the behavior prediction model includes a behavior time sub-model, and the inputting the multiple single-behavior time sequences into the behavior prediction model, so the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time includes:

inputting the multiple single-behavior time sequences into the behavior time sub-model in the behavior prediction model, and performing, by the behavior time sub-model, modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time.

6. The method according to claim 5, wherein the training the behavior prediction model based on the first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences includes:

training the behavior time sub-model in the behavior prediction model based on the first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences.

7. The method according to claim 5, wherein the behavior prediction model further includes a behavior relationship sub-model and a prediction sub-model, and the method further comprises:

inputting the multiple single-behavior sequences into the behavior relationship sub-model to obtain causality between every two single-behavior sequences in the multiple single-behavior sequences, wherein the causality is used to represent a relationship between occurrence of a behavior in one single-behavior sequence and occurrence of a behavior in another single-behavior sequence in every two single-behavior sequences;

inputting, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, and performing, by the prediction sub-model, sequence prediction based on the causality and the distribution situation to obtain a predicted behavior sequence; and

training the behavior relationship sub-model and the prediction sub-model based on the first loss value and a second loss value between the predicted behavior sequence and the behavior sequence of the user.

8. The method according to claim 7, wherein the inputting the multiple single-behavior sequences into the behavior relationship sub-model to obtain the causality between every two single-behavior sequences in the multiple single-behavior sequences includes:

obtaining the causality between every two single-behavior sequences in the multiple single-behavior sequences according to a sum of covariance values of every two single-behavior sequences in the multiple single-behavior sequences at all time points.

9. The method according to claim 7, wherein the training the behavior relationship sub-model and the prediction sub-model based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user includes:

determining a third loss value between a prediction label and a real label of the behavior sequence of the user, wherein the real label is used to indicate real information of the user, and the prediction label is used to indicate information of the user predicted by the prediction sub-model;

obtaining a fourth loss value by averaging the second loss value and the third loss value; and

training the behavior relationship sub-model and the prediction sub-model based on the first loss value and the fourth loss value.

10. The method according to claim 9, further comprising:

before the determining the third loss value between the prediction label and the real label of the behavior sequence of the user,

5 inputting, into the prediction sub-model, the causality between every two single-behavior sequences in the multiple single-behavior sequences and the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time, to obtain the prediction label corresponding to the behavior sequence of the user.

10 11. The method according to any one of claims 7 to 10, wherein the training the behavior relationship sub-model and the prediction sub-model based on the first loss value and the second loss value between the predicted behavior sequence and the behavior sequence of the user includes:

obtaining a fifth loss value according to the first loss value and the second loss value; and

15 training the behavior relationship sub-model based on the fifth loss value, and training the prediction sub-model based on the second loss value.

12. The method according to claim 11, wherein the obtaining the fifth loss value according to the first loss value and the second loss value includes:

20 determining a first value according to a matrix formed by a second-order partial derivation of each weight in the relationship sub-model and a vector formed by each weight in the relationship sub-model; and

obtaining the fifth loss value according to the first value, the first loss value, and the second loss value.

25 13. The method according to claim 1, further comprising:

inputting a behavior sequence to be predicted into the behavior prediction model to obtain a time point and a type corresponding to at least one future behavior of the behavior sequence to be predicted as predicted by the behavior prediction model.

30 14. An apparatus for training a behavior prediction model, comprising:

a determining module, configured to split a behavior sequence of a user to obtain multiple

single-behavior sequences, each single-behavior sequence being corresponding to one behavior, and being used to record a correspondence between the behavior and a time point;

the determining module being further configured to perform time coding on the multiple single-behavior sequences to obtain multiple single-behavior time sequences of the multiple single-behavior sequences;

the determining module being further configured to input the multiple single-behavior time sequences into a behavior prediction model, so the behavior prediction model performs modeling of the time point for each single-behavior time sequence in the multiple single-behavior time sequences to obtain a distribution situation of behaviors corresponding to the multiple single-behavior time sequences in terms of time; and

a training module, configured to train the behavior prediction model based on a first loss value between the distribution situation of the behaviors corresponding to the multiple single-behavior time sequences in terms of time and the multiple single-behavior time sequences, the behavior prediction model being used to predict a time point and a type of at least one future behavior of the user based on the behavior sequence of the user.

15. An electronic device, comprising a memory, a processor, and a computer program stored in the memory and capable of running on the processor, wherein when the processor executes the computer program, the electronic device is enabled to perform the method for training a behavior prediction model according to any one of claims 1 to 13.

16. A computer-readable storage medium, wherein the computer-readable storage medium stores instructions, and when the instructions run on a computer or a processor, the computer or the processor is enabled to perform the method for training a behavior prediction model according to any one of claims 1 to 13.

17. A computer program product comprising instructions, wherein when the computer program product runs on a computer or a processor, the computer or the processor is enabled to perform the method for training a behavior prediction model according to any one of claims 1 to 13.