

CLAIMS

What is claimed is:

1. A computer-implemented method based on deep reinforcement learning, comprising:
obtaining state information of an execution environment of a deep reinforcement learning system, wherein the execution environment comprises a referral marketing environment, and wherein the state information of the execution environment comprises user information of a current user and information associated with a current market of the deep reinforcement learning system, wherein the user information comprises user basic attribute information and user historical behavior information, and wherein the information associated with the current market comprises marketing services and transactions associated with the current user;

determining, by the deep reinforcement learning system, a marketing behavior corresponding to the state information of the execution environment using a marketing strategy that is trained by a deep neural network in the deep reinforcement learning system and has classified offers, wherein the marketing behavior comprises a combination of a marketing channel, marketing content, and a marketing time period, and wherein determining the marketing behavior comprises:

obtaining the user historical behavior information from the user information of the current user, wherein the user historical behavior information comprises at least one of a historical channel, a historical marketing content, or a historical transaction record;

generating a next marketing behavior using the user historical behavior information of the current user based on the classified offers in the marketing strategy, wherein the next marketing behavior comprises the combination of the marketing channel, the marketing content, and the marketing time period corresponding to the at least one of a historical channel, a historical marketing content, or a historical transaction record;

selecting the marketing channel, the marketing content, and the marketing time period from the next marketing behavior applicable to the execution environment; and

determining the marketing behavior using the selected marketing channel, the marketing content, and the marketing time period;

adjusting the execution environment by updating the marketing behavior;

in response to adjusting the execution environment by updating the marketing behavior, obtaining a current reward score of the execution environment for the marketing behavior, wherein the current reward score is determined based on a response of the current user to the marketing behavior; and

updating the marketing strategy based on the current reward score.

2. The method according to claim 1, wherein the user historical behavior information comprises at least one of the following: a historical operation sequence, a list of recently browsed pages, a list of recently accepted marketing information, a consumption record, a transaction record, a loan record, a travel record, and an insurance record.

3. The computer-implemented method according to claim 1, wherein the user information comprises user profile information predetermined based on at least one of the user basic attribute information or the user historical behavior information.

4. The computer-implemented method according to claim 1, wherein the state information further comprises scenario information of a scenario that the current user is located.

5. The computer-implemented method according to claim 3, wherein the scenario information comprises location service scenario information, payment scenario information, and wallet operation event information.

6. The computer-implemented method according to claim 1, wherein the marketing content comprises a marketing message and a credit package, and the credit package comprises one or more of a red envelope, a coupon, or a right.

7. The computer-implemented method according to claim 1, wherein the marketing channel comprises one or more of message sending, card display, badges reminder, and notification dots.

8. The computer-implemented method according to claim 1, further comprising:

determining a return value based on the current reward score, wherein the return value comprises a weighted sum of the current reward score and at least one future reward score; and
determining a subsequent marketing strategy to maximize the return value.

9. A non-transitory, computer-readable medium storing one or more instructions executable by a computer system to perform operations comprising:

obtaining state information of an execution environment of a deep reinforcement learning system, wherein the execution environment comprises a referral marketing environment, and wherein the state information of the execution environment comprises user information of a current user and information associated with a current market of the deep reinforcement learning system, wherein the user information comprises user basic attribute information and user historical behavior information, and wherein the information associated with the current market comprises marketing services and transactions associated with the current user;

determining, by the deep reinforcement learning system, a marketing behavior corresponding to the state information of the execution environment using a marketing strategy that is trained by a deep neural network in the deep reinforcement learning system and has classified offers, wherein the marketing behavior comprises a combination of a marketing channel, marketing content, and a marketing time period, and wherein determining the marketing behavior comprises:

obtaining the user historical behavior information of the current user from the user information of the current user, wherein the user historical behavior information comprises at least one of a historical channel, a historical marketing content, or a historical transaction record;

generating a next marketing behavior using the user historical behavior information of the current user based on the classified offers in the marketing strategy, wherein the next marketing behavior comprises the combination of the marketing channel, the marketing content, and the marketing time period corresponding to the at least one of a historical channel, a historical marketing content, or a historical transaction record;

selecting the marketing channel, the marketing content, and the marketing time period from the next marketing behavior applicable to the execution environment; and

determining the marketing behavior using the selected marketing channel, the marketing content, and the marketing time period;

adjusting the execution environment by updating the marketing behavior;
in response to adjusting the execution environment by updating the marketing behavior,
obtaining a current reward score of the execution environment for the marketing behavior, wherein
the current reward score is determined based on a response of the current user to the marketing
behavior; and
updating the marketing strategy based on the current reward score.

10. The apparatus according to claim 9, wherein the user historical behavior information
comprises at least one of the following: a historical operation sequence, a list of recently browsed
pages, a list of recently accepted marketing information, a consumption record, a transaction
record, a loan record, a travel record, and an insurance record.

11. The non-transitory, computer-readable medium according to claim 9, wherein the user
information comprises user profile information predetermined based on at least one of the user
basic attribute information or the user historical behavior information.

12. The non-transitory, computer-readable medium according to claim 9, wherein the state
information further comprises scenario information of a scenario that the current user is located.

13. The non-transitory, computer-readable medium according to claim 12, wherein the
scenario information comprises location service scenario information, payment scenario
information, and wallet operation event information.

14. The non-transitory, computer-readable medium according to claim 9, wherein the
marketing content comprises a marketing message and a credit package, and the credit package
comprises one or more of a red envelope, a coupon, or a right.

15. The non-transitory, computer-readable medium according to claim 9, wherein the
marketing channel comprises one or more of message sending, card display, badges reminder, and
notification dots.

16. The non-transitory, computer-readable medium according to claim 9, the operations further comprising:

determining a return value based on the current reward score, wherein the return value comprises a weighted sum of the current reward score and at least one future reward score; and
determining a subsequent marketing strategy to maximize the return value.

17. A computer-implemented system, comprising:

one or more computers; and

one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing one or more instructions that, when executed by the one or more computers, perform one or more operations comprising:

obtaining state information of an execution environment of a deep reinforcement learning system, wherein the execution environment comprises a referral marketing environment, and wherein the state information of the execution environment comprises user information of a current user and information associated with a current market of the deep reinforcement learning system, wherein the user information comprises user basic attribute information and user historical behavior information, and wherein the information associated with the current market comprises marketing services and transactions associated with the current user;

determining, by the deep reinforcement learning system, a marketing behavior corresponding to the state information of the execution environment using a marketing strategy that is trained by a deep neural network in the deep reinforcement learning system and has classified offers, wherein the marketing behavior comprises a combination of a marketing channel, marketing content, and a marketing time period, and wherein determining the marketing behavior comprises:

obtaining the user historical behavior information of the current user from the user information of the current user, wherein the user historical behavior information comprises at least one of a historical channel, a historical marketing content, or a historical transaction record;

generating a next marketing behavior using the user historical behavior information of the current user based on the classified offers in the marketing strategy, wherein the next marketing behavior comprises the combination of the marketing channel, the marketing

content, and the marketing time period corresponding to the at least one of a historical channel, a historical marketing content, or a historical transaction record;

selecting the marketing channel, the marketing content, and the marketing time period from the next marketing behavior applicable to the execution environment; and

determining the marketing behavior using the selected marketing channel, the marketing content, and the marketing time period;

adjusting the execution environment by updating the marketing behavior;

in response to adjusting the execution environment by updating the marketing behavior, obtaining a current reward score of the execution environment for the marketing behavior, wherein the current reward score is determined based on a response of the current user to the marketing behavior; and

updating the marketing strategy based on the current reward score.

18. The computer-implemented system according to claim 17, wherein the user information comprises user basic attribute information.

19. The apparatus according to claim 17, wherein the user historical behavior information comprises at least one of the following: a historical operation sequence, a list of recently browsed pages, a list of recently accepted marketing information, a consumption record, a transaction record, a loan record, a travel record, and an insurance record.

20. The computer-implemented system according to claim 17, wherein the state information further comprises scenario information of a scenario that the current user is located.

21. The computer-implemented system according to claim 20, wherein the scenario information comprises location service scenario information, payment scenario information, and wallet operation event information.

22. The computer-implemented system according to claim 17, wherein the marketing content comprises a marketing message and a credit package, and the credit package comprises one or more of a red envelope, a coupon, or a right.

23. The computer-implemented system according to claim 17, wherein the marketing channel comprises one or more of message sending, card display, badge reminder, and notification dots.

24. The computer-implemented system according to claim 17, the operations further comprising:

determining a return value based on the current reward score, wherein the return value comprises a weighted sum of the current reward score and at least one future reward score; and
determining a subsequent marketing strategy to maximize the return value.