

CLAIMS

What is claimed is:

1. A marketing method based on a deep reinforcement learning system, wherein the deep reinforcement learning system comprises an agent and an execution environment, the agent is configured to determine a marketing behavior based on a marketing strategy and status information of the execution environment, and the method comprises:

obtaining, from the execution environment, multiple execution results generated when a user responds to the marketing behavior, wherein the multiple execution results respectively correspond to multiple target effects on a marketing effect chain;

determining a reward score of reinforcement learning based on the multiple execution results by:

determining at least one independent function f_i and least one combination function F_i , wherein the independent function f comprises at least one execution result (X_i) selected from the multiple execution results, and the combination function F comprises at least two execution results (X_i, X_{i+1}) selected from the multiple execution results;

determining a target effect Y_i for each execution result X_i , wherein the execution result X_i is designed to be closer to the target effect Y_i , and wherein the independent function f_i and the combination function F_i are designed to output a higher score when the execution result X_i is closer to the target effect Y_i ;

determining the reward score based on a score evaluation function:

$$R_t = W_1 \times f_1(X_1) + W_2 F_2 \times F_2(X_2, X_3) \dots$$

wherein R_t is the reward score and W_i is a weight predetermined corresponding to each execution result X_i ; and

returning the reward score to the agent for the agent to update its marketing strategy.

2. The method according to claim 1, wherein the obtaining, from the execution environment, multiple execution results generated when a user responds to the marketing behavior comprises:

obtaining an interaction behavior of the user from a server, and obtaining the multiple execution results by using the interaction behavior; or

capturing an operation behavior of the user by using a client device user interface

tracing point, and obtaining the multiple execution results by using the operation behavior of the user.

3. The method according to claim 1, wherein the multiple target effects on the marketing effect chain comprise multiple targets in the following: reach, a first click,
5 stay, a further click, and conversion of marketing information.

4. The method according to claim 1, wherein the determining a reward score of reinforcement learning comprises determining the reward score by using the score evaluation function, wherein the score evaluation function uses the multiple execution results as variables and is designed to be positively correlated with a match rate between
10 the multiple execution results and multiple corresponding target effects.

5. The method according to claim 1, wherein the determining a reward score of reinforcement learning comprises:

obtaining independent functions constructed separately for at least one of the multiple execution results, and determining the reward score based on at least a
15 weighted sum of the independent functions.

6. The method according to claim 5, wherein at least one of the independent functions is constructed to be positively correlated with a corresponding match rate between the execution result and the target effect.

7. The method according to claim 5, wherein the multiple execution results
20 comprise a first execution result, and the first execution result has a discrete result value; and the independent functions comprise a first function, and the first function outputs a discrete function value based on the discrete result value of the first execution result.

8. The method according to claim 7, wherein the first function outputs a first discrete value when the first execution result has a first result value and outputs a second
25 discrete value when the first execution result has a second result value, wherein the first result value has a higher match rate with a first target effect corresponding to the first execution result than the second result value, and the first discrete value is greater than the second discrete value.

9. The method according to claim 1 or any one of claims 5 to 8, wherein the
30 determining a reward score of reinforcement learning comprises:

determining at least one result combination based on the multiple execution results;
obtaining the at least one combination function constructed for the at least one result combination; and

determining the reward score based on at least a weighted sum of the at least one

combination function.

10. The method according to claim 1, wherein the determining a reward score of reinforcement learning comprises:

5 determining the target effect or a target effect combination that matches the multiple execution results;

obtaining a pre-established mapping relationship between the target effect or a target effect combination and the reward score; and

determining, based on the mapping relationship, a reward score corresponding to the matched target effect or target effect combination.

10 11. A marketing apparatus based on a deep reinforcement learning system, wherein the deep reinforcement learning system comprises an agent and an execution environment, the agent is configured to determine a marketing behavior based on a marketing strategy and status information of the execution environment, and the apparatus comprises:

15 an acquisition unit, configured to obtain, from the execution environment, multiple execution results generated when a user responds to the marketing behavior, wherein the multiple execution results respectively correspond to multiple target effects on a marketing effect chain;

20 a determining unit, configured to determine a reward score of reinforcement learning based on the multiple execution results by:

25 determining at least one independent function f_i and least one combination function F_i , wherein the independent function f comprises at least one execution result (X_i) selected from the multiple execution results, and the combination function F comprises at least two execution results (X_i, X_{i+1}) selected from the multiple execution results;

determining a target effect Y_i for each execution result X_i , wherein the execution result X_i is designed to be closer to the target effect Y_i , and wherein the independent function f_i and the combination function F_i are designed to output a higher score when the execution result X_i is closer to the target effect Y_i ;

30 determining the reward score based on a score evaluation function:

$$R_t = W_1 \times f_1(X_1) + W_2 F_2 \times F_2(X_2, X_3) \dots$$

wherein R_t is the reward score and W_i is a weight predetermined corresponding to each execution result X_i ; and

a returning unit, configured to return the reward score to the agent for the agent to

update its marketing strategy.

12. The apparatus according to claim 11, wherein the acquisition unit is configured to:

5 obtain an interaction behavior of the user from a server, and obtain the multiple execution results by using the interaction behavior; or

capture an operation behavior of the user by using a client device user interface tracing point, and obtain the multiple execution results by using the operation behavior of the user.

10 13. The apparatus according to claim 11, wherein the multiple target effects on the marketing effect chain comprise multiple targets in the following: reach, a first click, stay, a further click, and conversion of marketing information.

14. The apparatus according to claim 11, wherein the determining unit is configured to determine the reward score by using the score evaluation function, wherein the score evaluation function uses the multiple execution results as variables
15 and is designed to be positively correlated with a match rate between the multiple execution results and multiple corresponding target effects.

15. The apparatus according to claim 11, wherein the determining unit is configured to:

20 obtain independent functions constructed separately for at least one of the multiple execution results, and determine the reward score based on at least a weighted sum of the independent functions.

16. The apparatus according to claim 15, wherein at least one of the independent functions is constructed to be positively correlated with a corresponding match rate between the execution result and the target effect.

25 17. The apparatus according to claim 15, wherein the multiple execution results comprise a first execution result, and the first execution result has a discrete result value; and the independent functions comprise a first function, and the first function outputs a discrete function value based on the discrete result value of the first execution result.

30 18. The apparatus according to claim 17, wherein the first function outputs a first discrete value when the first execution result has a first result value and outputs a second discrete value when the first execution result has a second result value, wherein the first result value has a higher match rate with a first target effect corresponding to the first execution result than the second result value, and the first discrete value is greater than the second discrete value.

19. The apparatus according to claim 11 or any one of claims 15 to 18, wherein the determining unit is configured to:

determine at least one result combination based on the multiple execution results;

5 obtain the at least one combination function constructed for the at least one result combination; and

determine the reward score based on at least a weighted sum of the at least one combination function.

20. The apparatus according to claim 11, wherein the determining unit is configured to:

10 determine the target effect or a target effect combination that matches the multiple execution results;

obtain a pre-established mapping relationship between the target effect or a target effect combination and the reward score; and

15 determine, based on the mapping relationship, a reward score corresponding to the matched target effect or target effect combination.

21. A computing device, comprising a memory and a processor, wherein the memory stores executable code, and when executing the executable code, the processor implements the method according to any one of claims 1 to 10.