CLAIMS

What is claimed is:

1. A method comprising:

obtaining a plurality of candidate page display methods for displaying a succession of pages;

obtaining a plurality of candidate page element arrangement methods for arranging different page elements on a same page;

obtaining a plurality of candidate attributes of the page elements;

generating, from the candidate page display methods, the candidate page element arrangement methods, and the candidate attributes of the page elements, a first plurality of candidate page layouts;

generating a display policy that defines a mapping relationship between the first plurality of candidate page layouts and a plurality of different user features;

in response to a request to access a first page by a first user, obtaining a plurality of first user features, wherein the plurality of first user features comprises one or more user profile features and one or more user context features;

selecting, according to the display policy and based on the plurality of first user features, a first page layout from the first plurality of candidate page layouts;

displaying the first page with the first page layout;

obtaining operation information of the first user interacting with the first page with the first page layout, and determining a reward score based on the operation information; and

updating the display policy by implementing a reinforcement learning algorithm based at least on the reward score.

2. The method according to claim 1, wherein

the user profile features comprise at least one of the following: an age, a gender, registration duration, an education level, an occupation, or a user tag; and

the user context features comprise at least one of the following: time when the request is sent, an interface that the request is sent through, or operation track information of the first user.

3. The method according to claim 1, wherein

the candidate page display methods comprise waterfall display or page turning display; and

the plurality of candidate attributes of the page elements comprise display sizes, display shapes, or resolution of the page elements.

4. The method according to claim 1, wherein the display policy is an initial policy, and selecting the first page layout from the first plurality of candidate page layouts generated for the first page comprises:

randomly assigning a first traffic bucket in a first plurality of traffic buckets to the first user, wherein the first plurality of traffic buckets correspond to the first plurality of candidate page layouts; and

using a candidate page layout from the first plurality of candidate page layouts that corresponds to the first traffic bucket as the first page layout.

5. The method according to claim 1, wherein the display policy is implemented by one of a logistic regression model, a gradient boosting decision tree GBDT model, or a deep neural network.

6. The method according to claim 1, wherein the operation information comprises at least one of the following: hit count on the first page, first page view count, view depth to the first page, or time on the first page; and the determining the reward score for current page display based on the operation information comprises: determining the reward score as a weighted sum of at least one of the operation information.

7. The method according to claim 1, wherein the operation information comprises a transaction amount achieved based on the first page; and the determining the reward score for current page display based on the operation information comprises: determining the reward score to be directly proportional to the transaction amount.

8. The method according to claim 1, wherein updating the display policy for displaying pages by using the reinforcement learning algorithm based at least on the reward score to perform page display comprises:

adding the first user features, the first page layout, and the reward score to a training sample set as one training sample; and

updating the display policy for displaying the pages on the training sample set and by using the reinforcement learning algorithm.

9. The method according to claim 1, wherein the reinforcement learning algorithm comprises a SARSA algorithm, an evolution strategy algorithm, a Q-learning algorithm, a policy gradient algorithm, or a cross information entropy method.

10. 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 operations comprising:

obtaining a plurality of candidate page display methods for displaying a succession of pages;

obtaining a plurality of candidate page element arrangement methods for arranging different page elements on a same page;

obtaining a plurality of candidate attributes of the page elements;

generating, from the candidate page display methods, the candidate page element arrangement methods, and the candidate attributes of the page elements, a first plurality of candidate page layouts;

generating a display policy that defines a mapping relationship between the first plurality of candidate page layouts and a plurality of different user features;

in response to a request to access a first page by a first user, obtaining a plurality of first user features, wherein the plurality of first user features comprises one or more user profile features and one or more user context features;

selecting, according to the display policy and based on the plurality of first user features, a first page layout from the first plurality of candidate page layouts;

displaying the first page with the first page layout;

obtaining operation information of the first user interacting with the first page with the first page layout, and determining a reward score based on the operation information; and

updating the display policy by implementing a reinforcement learning algorithm based at least on the reward score.

11. The computer-implemented system according to claim 10, wherein

the user profile features comprise at least one of the following: an age, a gender, registration duration, an education level, an occupation, or a user tag; and

the user context features comprise at least one of the following: time when the request is sent, an interface that the request is sent through, or operation track information of the first user.

12. The computer-implemented system according to claim 10, wherein

the candidate page display methods comprise waterfall display or page turning display; and

the plurality of candidate attributes of the page elements comprise display sizes, display shapes, or resolution of the page elements.

13. The computer-implemented system according to claim 10, wherein the display policy is an initial policy, and selecting the first page layout from the first plurality of candidate page layouts generated for the first page comprises:

randomly assigning a first traffic bucket in a first plurality of traffic buckets to the first user, wherein the first plurality of traffic buckets correspond to the first plurality of candidate page layouts; and

using a candidate page layout from the first plurality of candidate page layouts that corresponds to the first traffic bucket as the first page layout.

14. The computer-implemented system according to claim 10, wherein the display policy is implemented by one of a logistic regression model, a gradient boosting decision tree GBDT model, or a deep neural network.

15. The computer-implemented system according to claim 10, wherein the operation information comprises at least one of the following: hit count on the first page, first page view count, view depth to the first page, or time on the first page; and

the determining the reward score for current page display based on the operation information comprises: determining the reward score as a weighted sum of at least one of the operation information.

16. The computer-implemented system according to claim 10, wherein the operation information comprises a transaction amount achieved based on the first page; and

the determining the reward score for current page display based on the operation information comprises: determining the reward score to be directly proportional to the transaction amount.

17. The computer-implemented system according to claim 10, wherein updating the display policy for displaying pages by using the reinforcement learning algorithm based at least on the reward score to perform page display comprises:

adding the first user features, the first page layout, and the reward score to a training sample set as one training sample; and

updating the display policy for displaying pages on the training sample set and by using the reinforcement learning algorithm.

18. The computer-implemented system according to claim 10, wherein the reinforcement learning algorithm comprises an evolution strategy algorithm, a Q-learning algorithm, a SARSA algorithm, a policy gradient algorithm, or a cross information entropy method.

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

obtaining a plurality of candidate page display methods for displaying a succession of pages;

obtaining a plurality of candidate page element arrangement methods for arranging different page elements on a same page;

obtaining a plurality of candidate attributes of the page elements;

generating, from the candidate page display methods, the candidate page element arrangement methods, and the candidate attributes of the page elements, a first plurality of candidate page layouts;

generating a display policy that defines a mapping relationship between the first plurality of candidate page layouts and a plurality of different user features;

in response to a request to access a first page by a first user, obtaining a plurality of first user features, wherein the plurality of first user features comprises one or more user profile features and one or more user context features;

selecting, according to the display policy and based on the plurality of first user features, a first page layout from the first plurality of candidate page layouts;

displaying the first page with the first page layout;

obtaining operation information of the first user interacting with the first page with the first page layout, and determining a reward score based on the operation information; and

updating the display policy by implementing a reinforcement learning algorithm based at least on the reward score.