

CLAIMS

1. A voice packet recommendation method, comprising:

selecting at least one target display video for a user from among a plurality of candidate display videos associated with a plurality of voice packets and using voice packets to which the at least one target display video belongs as candidate voice packets;

selecting a target voice packet for the user from among the candidate voice packets according to attribute information of the candidate voice packets and attribute information of the at least one target display video; and

recommending the target voice packet to the user.

2. The method of claim 1, wherein the selecting at least one target display video for a user from among a plurality of candidate display videos associated with a plurality of voice packets comprises:

determining the at least one target display video according to a degree of relevance between a portrait tag of the user and a plurality of classification tags of the candidate display videos associated with the voice packets.

3. The method of claim 2, further comprising:

extracting a plurality of pictures from each of the candidate display videos; and

inputting the extracted pictures into a pretrained multi-classification model and determining at least one classification tag of the each of the candidate display videos according to a model output result.

4. The method of claim 3, further comprising:

using a text description of a sample video, or a user portrait of a viewing user of a sample video, or a text description of a sample video and a user portrait of a viewing user of the sample video as a sample classification tag of the sample video; and

training a preconstructed neural network model according to a sample picture extracted from the sample video and the sample classification tag to obtain the multi-classification model.

5. The method of claim 3, wherein the multi-classification model shares model parameters in a process of determination of each of the classification tags.

6. The method of claim 2, wherein each of the classification tags comprises at least one of an image tag, a voice quality tag or a voice style tag.

7. The method of claim 1, further comprising:

determining initial display videos of each of the voice packets; and

5 determining, according to a video source priority level of each of the initial display videos, candidate display videos associated with the each of the voice packets.

8. The method of claim 1, further comprising:

determining initial display videos of each of the voice packets; and

10 determining, according to similarity between each of the initial display videos and the each of the voice packets, candidate display videos associated with the each of the voice packets.

9. The method of claim 7 or 8, wherein the determining initial display videos of each of the voice packets comprises:

determining promotion text of the each of the voice packets according to a promotion picture of a provider of the each of the voice packets;

15 generating a promotion audio and a promotion caption according to the promotion text based on an acoustic synthesis model of the provider of the each of the voice packets; and

generating the initial display videos according to the promotion picture, the promotion audio and the promotion caption.

20 10. The method of claim 7 or 8, wherein the determining initial display videos of each of the voice packets comprises:

constructing a video search word according to information about a provider of the each of the voice packets; and

25 searching for videos of the provider of the each of the voice packets according to the video search word and using the videos of the provider of the each of the voice packets as the initial display videos.

11. The method of any one of claims 1 to 8, wherein the recommending the target voice packet to the user comprises:

recommending the target voice packet to the user through a target display video associated with the target voice packet.

12. A voice packet recommendation apparatus, comprising:

5 a target display video selection module configured to select at least one target display video for a user from among a plurality of candidate display videos associated with a plurality of voice packets and use voice packets to which the at least one target display video belongs as candidate voice packets;

10 a target voice packet selection module configured to select a target voice packet for the user from among the candidate voice packets according to attribute information of the candidate voice packets and attribute information of the at least one target display video; and

a target voice packet recommendation module configured to recommend the target voice packet to the user.

13. The apparatus of claim 12, wherein the target display video selection module comprises:

15 a target display video determination unit configured to determine the at least one target display video according to a degree of relevance between a portrait tag of the user and classification tags of the candidate display videos associated with the voice packets.

14. The apparatus of claim 13, further comprising:

a picture extraction module configured to extract a plurality of pictures from each of the candidate display videos; and

20 a classification tag determination module configured to input the extracted pictures into a pretrained multi-classification model and determine at least one classification tag of the each of the candidate display videos according to a model output result.

15. The apparatus of claim 14, further comprising:

25 a sample classification tag determination module configured to use a text description of a sample video, or a user portrait of a viewing user of a sample video, or a text description of a sample video and a user portrait of a viewing user of the sample video as a sample classification tag of the sample video; and

a multi-classification model training module configured to train a preconstructed neural network model according to a sample picture extracted from the sample video and the sample classification

tag to obtain the multi-classification model.

16. The apparatus of claim 14, wherein the multi-classification model shares model parameters in a process of determination of each of the classification tags.

5 17. The apparatus of claim 13, wherein each of the classification tags comprises at least one of an image tag, a voice quality tag or a voice style tag.

18. The apparatus of claim 12, further comprising:

an initial display video determination module configured to determine initial display videos of each of the voice packets; and

10 a candidate display video determination module configured to determine, according to a video source priority level of each of the candidate display videos, candidate display videos associated with the each of the voice packets.

19. An electronic device, comprising:

at least one processor; and

a memory which is in communication connection to the at least one processor, wherein

15 the memory stores instructions executable by the at least one processor, wherein the instructions are configured to, when executed by at least one processor, cause the at least one processor to perform the voice packet recommendation method of any one of claims 1 to 11.

20 20. A non-transitory computer-readable storage medium storing computer instructions, wherein the computer instructions are configured to cause a computer to perform the voice packet recommendation method of any one of claims 1 to 11.