

CLAIMS

1. A voice packet recommendation method, comprising:

5 constructing a first video training sample according to first user behavior data of a first sample user in a video recommendation scenario and first video data associated with the first user behavior data;

constructing a user training sample according to sample search data of the first sample user and historical interaction data about a first sample voice packet;

pretraining a neural network model according to the first video training sample and the user training sample; and

10 retraining the pretrained neural network model by using a sample video and sample tag data which are associated with a second sample voice packet to obtain a voice packet recommendation model.

2. The method of claim 1, further comprising:

training a preconstructed video feature vector representation network; and

15 constructing the neural network model according to the trained video feature vector representation network.

3. The method of claim 2, wherein the "training a preconstructed video feature vector representation network" comprises:

20 constructing a second video training sample from second user behavior data of a second sample user in the video recommendation scenario and second video data associated with the second user behavior data; and

training the preconstructed video feature vector representation network according to the second video training sample.

25 4. The method of claim 1, wherein the "retraining the pretrained neural network model by using a sample video and sample tag data which are associated with a second sample voice packet" comprises:

inputting the sample video and the sample tag data to the pretrained neural network model to

adjust network parameters of a fully-connected layer in the pretrained neural network model.

5. The method of claim 1, further comprising:

determining candidate sample videos of the second sample voice packet; and

5 determining, according to a video source priority level of each of the candidate sample videos, the sample video associated with the second sample voice packet.

6. The method of claim 1, further comprising:

determining candidate sample videos of the second sample voice packet; and

determining, according to similarity between each of the candidate sample videos and the second sample voice packet, the sample video associated with the second sample voice packet.

10 7. The method of claim 5 or 6, wherein the "determining candidate sample videos of the second sample voice packet" comprises:

determining promotion text of the second sample voice packet according to a promotion picture of a voice packet provider of the second sample voice packet;

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

generating the candidate sample videos according to the promotion picture, the promotion audio and the promotion caption.

8. The method of claim 5 or 6, wherein the "determining candidate sample videos of the second sample voice packet" comprises:

20 constructing a video search word according to information about a voice packet provider of the second sample voice packet; and

searching for videos of the voice packet provider according to the video search word and using the videos of the voice packet provider as the candidate sample videos.

9. The method of any one of claims 1 to 6, further comprising:

25 inputting each candidate display video of a user for recommendation, description text of the each candidate display video, a historical search word of the user, and a historical voice packet

used by the user to the voice packet recommendation model; and

recommending a target display video comprising download information of a target voice packet to the user for recommendation according to a model output result of the voice packet recommendation model.

- 5 10. The method of any one of claims 1 to 6, wherein the first user behavior data comprises behavior data about user behavior of browsing completed, upvoting and adding to favorites; the first video data comprises video content and description text of a first video associated with the first user behavior data; and the historical interaction data is voice packet usage data.

11. A voice packet recommendation apparatus, comprising:

- 10 a first video training sample construction module configured to construct a first video training sample according to first user behavior data of a first sample user in a video recommendation scenario and first video data associated with the first user behavior data;

15 a user training sample construction module configured to construct a user training sample according to sample search data of the first sample user and historical interaction data about a first sample voice packet;

15 a pretraining module configured to pretrain a neural network model according to the first video training sample and the user training sample; and

- 20 a retraining module configured to retrain the pretrained neural network model by using a sample video and sample tag data which are associated with a second sample voice packet to obtain a voice packet recommendation model.

12. The apparatus of claim 11, further comprising:

a video feature vector representation network training module configured to train a preconstructed video feature vector representation network; and

- 25 a neural network model construction module configured to construct the neural network model according to the trained video feature vector representation network.

13. The apparatus of claim 12, wherein the video feature vector representation network training module comprises:

a second video training sample construction unit configured to construct a second video training

sample from second user behavior data of a second sample user in the video recommendation scenario and second video data associated with the second user behavior data; and

a video feature vector representation network training unit configured to train the preconstructed video feature vector representation network according to the second video training sample.

5 14. The apparatus of claim 11, wherein the retraining module comprises:

a network parameter adjustment unit configured to input the sample video and the sample tag data to the pretrained neural network model to adjust network parameters of a fully-connected layer in the neural network model.

15. The apparatus of claim 11, further comprising:

10 a candidate sample video determination module configured to determine candidate sample videos of the second sample voice packet; and

a sample video determination module configured to determine, according to a video source priority level of each of the candidate sample videos, the sample video associated with the second sample voice packet.

15 16. The apparatus of claim 11, further comprising:

a candidate sample video determination module configured to determine candidate sample videos of the second sample voice packet; and

20 a sample video determination module configured to determine, according to similarity between each of the candidate sample videos and the second sample voice packet, the sample video associated with the second sample voice packet.

17. The apparatus of claim 15 or 16, wherein the candidate sample video determination module comprises:

25 a promotion text determination unit configured to determine promotion text of the second sample voice packet according to a promotion picture of a voice packet provider of the second sample voice packet;

an audio and caption determination unit configured to generate a promotion audio and a promotion caption according to the promotion text based on an acoustic synthesis model of the voice packet provider; and

a candidate sample video generation unit configured to generate the candidate sample videos according to the promotion picture, the promotion audio and the promotion caption.

18. The apparatus of claim 15 or 16, wherein the candidate sample video determination module comprises:

5 a video search word construction unit configured to construct a video search word according to information about a voice packet provider of the second sample voice packet; and

a candidate sample video determination unit configured to search for videos of the voice packet provider according to the video search word and use the videos of the voice packet provider as the candidate sample videos.

10 19. An electronic device, comprising:

at least one processor; and

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

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
15 perform the voice packet recommendation method of any one of claims 1 to 10.

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 10.