CLAIMS

1. A computer-implemented method for personalized speaker verification, comprising:

at a first stage of training a first model for determining voice characteristics;

obtaining first speech data of a speaker as a positive sample and second speech data of an entity different from the speaker as a negative sample;

feeding the positive sample and the negative sample to the first model for determining voice characteristics to correspondingly output a positive voice characteristic and a negative voice characteristic of the speaker, wherein the first model has not been personalized for speaker verification;

generating a template vector of the speaker based on the positive voice characteristic;

obtaining a gradient based at least on the positive voice characteristic and the negative voice characteristic;

at a second stage of training the first model for determining voice characteristics, feeding the gradient to the first model to update one or more parameters of the first model to obtain a second model for personalized speaker verification;

obtaining speech data of a user;

feeding the obtained speech data to the second model to generate an input vector of the user;

comparing the input vector of the user with the template vector of the speaker; and

verifying if the user is the speaker based at least on the comparison.

2. The method of claim 1, wherein:

obtaining the first speech data of the speaker as the positive sample and the second speech data of the entity different from the speaker as the negative sample comprises: obtaining one or more speech segments of the speaker as the positive sample and one or more speech segments of one or more people other than the speaker as the negative sample; and

feeding the positive sample and the negative sample to the first model to correspondingly output the positive voice characteristic and the negative voice characteristic of the speaker comprises: feeding the one or more speech segments of the speaker into the first model to correspondingly output one or more positive sample vectors, and feeding the one or more speech

segments of the one or more people other than the speaker into the first model to correspondingly output one or more negative sample vectors.

3. The method of claim 1, wherein generating the template vector of the speaker based on the positive voice characteristic comprises:

averaging the one or more positive sample vectors to obtain the template vector of the speaker.

4. The method of claim 2, wherein obtaining the gradient based at least on the positive voice characteristic and the negative voice characteristic comprises:

feeding the one or more positive sample vectors and the one or more negative sample vectors into a neural network classifier to obtain one or more gradient vectors.

5. The method of claim 4, wherein obtaining the gradient based at least on the positive voice characteristic and the negative voice characteristic further comprises:

averaging the one or more gradient vectors to obtain an average gradient vector of the speaker as the gradient.

6. The method of claim 5, wherein:

feeding the gradient to the first model to update the one or more parameters of the first model comprises: feeding the average gradient vector of the speaker to the first model to update the one or more parameters of the first model; and

the one or more parameters associate different neural layers of the first model.

7. The method of claim 4, wherein feeding the one or more positive sample vectors and the one or more negative sample vectors into the neural network classifier to obtain one or more gradient vectors comprises:

obtaining the gradient based at least on backward propagation through a cross-entropy loss function of the neural network classifier.

8. The method of claim 1, wherein feeding the gradient to the first model to update the one or more parameters of the first model comprises:

feeding the gradient to the first model to update the one or more parameters of the first model based at least on the gradient and an online machine learning rate.

9. The method of claim 8, feeding the gradient to the first model to update the one or more parameters of the first model based at least on the gradient and the online machine learning rate comprises:

updating the one or more parameters in a direction in which the gradient descends at a fastest online machine learning rate.

10. The method of claim 1, wherein:

before feeding the positive sample and the negative sample to the first model for determining voice characteristics, the first model has been trained at least by jointly minimizing a first loss function that optimizes speaker classification and a second loss function that optimizes speaker clustering.

11. The method of claim 10, wherein:

the first loss function comprises a non-sampling-based loss function; and

the second function comprises a Gaussian mixture loss function with non-unit multi-variant covariance matrix.

12. A system for personalized speaker verification, comprising:

one or more processors; and

one or more computer-readable memories coupled to the one or more processors and having instructions stored thereon that are executable by the one or more processors to perform the method of any of claims 1-11.

13. An apparatus for personalized speaker verification comprising a plurality of modules for performing the method of any of claims 1-11.

14. A non-transitory computer-readable storage medium configured with instructions executable by one or more processors to cause the one or more processors to perform the method of any of claims 1-11.