SPEECH INTERACTION MODEL SYSTEMS AND METHODS

Abstract

The present disclosure relates to speech interaction models, and in particular to methods and systems for training a speech interaction model. The methods and systems comprise receiving a training sample set including a plurality of question texts, corresponding sample response texts for each question text, and question audios associated with the question texts, and inputting the training sample set into the speech interaction model to train the model. The speech interaction model is configured to generate question text features corresponding to each question text, generate predicted response texts, and generate question audio features corresponding to the question audios. The method further comprises determining a first loss value of the trained speech interaction model based on the question text features and the question audio features corresponding to each question text, determining a second loss value of the trained speech interaction model based on the predicted response texts and the sample response texts corresponding to each question text, determining if the trained speech interaction model has converged based on the first and second loss values, and establishing the trained speech interaction model if the trained speech interaction model is determined to be convergent based on the first and second loss values.

Fig. 1