A Deep Learning Method Using Gender-Specific Features for Emotion Recognition

ABSTRACT

Speech reflects people's mental state and using a microphone sensor is a potential method for human–computer interaction. Speech recognition using this sensor is conducive to the diagnosis of mental illnesses. The gender difference of speakers affects the process of speech emotion recognition based on specific acoustic features, resulting in the decline of emotion recognition accuracy. Therefore, we believe that the accuracy of speech emotion recognition can be effectively improved by selecting different features of speech for emotion recognition based on the speech representations of different genders. In this paper, we propose a speech emotion recognition method based on gender classification. First, we use MLP to classify the original speech by gender. Second, based on the different acoustic features of male and female speech, we analyze the influence weights of multiple speech emotion features in male and female speech, and establish the optimal feature sets for male and female emotion recognition, respectively. Finally, we train and test CNN and BiLSTM, respectively, by using the male and the female speech emotion feature sets. The results show that the proposed emotion recognition models have an advantage in terms of average recognition accuracy compared with gender-mixed recognition models