Speech Emotion Recognition-Based on Multi-Feature and Multi-Language Fusion and Its Application in Facial Expression Editing

Wang, Chunyi (School: The Experimental High School Attached to Beijing Normal University)

A speech emotion recognition algorithm based on multi-feature and multi-lingual fusion is proposed in order to resolve low recognition accuracy caused by lack of large speech dataset and low robustness of acoustic features in the recognition of speech emotion. First, handcrafted and deep automatic features are extracted from existing data in different languages. Then, the various features are fused respectively and trained in a classification model. Distinguishing the fused features with the unfused ones, the results manifest that the fused features significantly enhance the accuracy of speech emotion recognition algorithm. The proposed solution is evaluated on the two Chinese corpus and two English corpus, and is shown to provide more accurate predictions compared to original solution. As a result of this study, the multi-feature and Multi-lingual fusion algorithm can significantly improve the speech emotion recognition accuracy when the dataset is small. In addition, the GANimation model is used to generate facial images with special emotion, which is the result of speech emotion recognition. GANimation is a novel GAN (Generative Adversarial Network) conditioning scheme based on Action Units (AU) annotations, which describes in a continuous manifold anatomical facial movements defining a human expression. In this work, the result of speech emotion recognition is mapped to the corresponding input signal to generate facial emotion which closes to the speech. In engineering, in order to improve the efficiency of facial expression editing, we also try to prune and reduce model parameters to improve model prediction speed, and ultimately improve the effect speed of the application.

Awards Won: