Abstract

Affective computing, is a growing research area used to develop the system in such a way to recognize, interpret, process and simulate the human emotions in a systematic manner. The main application of Affective computing is the human computer interaction, in which the communication between the human and the machine enhances by giving an appropriate response to the user in an effective and empathic manner. The system can collect the emotion specific information from various sources like facial expressions, postures, gestures and speech etc. This thesis mainly concentrate on the systems which can extract the emotional specific information of the human through speech and proposes various methodologies in automatic identification of emotions. Speech signal is a natural and effective way of communication which contains mainly two different types of information. One is the explicit message by the speaker and the other is the implicit information about the speaker. Implicit information contains the information about speakers age, health condition, gender, and emotion of the speaker. This work focuses mainly on two concepts. One is ‘feature fusion technique’ and the other is ‘two stage emotion classification technique’.

In feature fusion technique, the most popular features in prosody and spectral properties of the speech signal are fused together. These fused features are classified using various classification techniques like Linear Discriminant Analysis(LDA), Regularized Discriminant Analysis(RDA), Support Vector Machine(SVM) and k Nearest Neighbor(kNN). The results are validated over Spanish and Berlin emotional speech databases. The results reveal that the performance of each classifier is improved effectively with feature fusion rather than that of individual features and also an effective comparison of results is done with each classifier using fused features and individual features. The comparison of results
showed that the RDA is a better classification technique, as it eliminates the sin-
gularity problem that occurs in LDA which is due to high dimensional and small
sample size problem.

To improve the performance of speech emotion recognition system fur-
ther, a ‘two stage feature selection’ technique is proposed. In the first stage feature
selection, appropriate features are selected and fused together. The emotions con-
sidered in this work are happy, neutral, anger, sad, fear and disgust from the
Berlin and Spanish emotional speech databases. Even though the performance
of the system is improved with this technique, it does not reach to an optimal
state because of high dimensional correlated feature set. To avoid this problem,
we need second stage feature selection. In this stage, different feature subset se-
lection techniques like Sequential Forward Selection (SFS), Sequential Floating
Forward Selection (SFBS) are used to get an optimal uncorrelated feature subset
with which the performance of the speech emotion recognition system is enhanced.

The extensive application areas of speech emotion recognition are e-
learning in which the learning methodologies depend on the response of the stu-
dent, entertainment in which the systems response depends on the users level of
thinking and in medical diagnostic tool the recognized emotions can help a speech
therapist in treating his/her patients. In this thesis, we are particularly interested
in Advanced Road Safety Driver Assistance System in which the safety of the
driver is given the highest priority by detecting the emotion of the driver to make
him alert from an accident. This is all done by detecting the emotion of the driver
through his speech samples.