ERCIM “Alain Bensoussan”
Fellowship Scientific Report

Fellow: Margarita Kotti
Visited Location : Human Interface in Information Systems (*HIIS*) Laboratory at Istituto di Scienza e Tecnologie dell'Informazione (*ISTI*) of National Research Council (*CNR*)
Duration of Visit: 01 October 2009 – 30 June 2010

I - Scientific activity

(1 page at maximum)

Our research has focused on emotion recognition of the vocal interaction. It is proven that when a computer ignores human emotional responses, users are likely to feel strong resistance towards continuing the interaction. On the contrary, a user interface is more likely to get accepted by the user if it is sensitive towards the user's feelings.

We propose an approach for emotion recognition that consists of three steps. Initially feature extraction is carried out, secondly, feature selection is applied, and finally classification is carried out. The initial system built is able to distinguish between negative and non-negative emotions. Extracted features are related to statistics of pitch, formant, and energy contours, as well as on autocorrelation, MPEG-7 descriptors, Fujisaki's model parameters, jitter, and shimmer. Feature selection is applied next by means of the Best First algorithm. The selected features are then fed as input to support vector machines with various kernel functions as well as to the K-nearest neighbors classifier. Speaker-independent leave-one-speaker-out evaluation is carried out on the Berlin emotional speech database. The best performing classifier is found to be the support vector machine with the Gaussian radial basis function kernel. Correctly classified utterances are 86.73% for male subjects and 91.73% for female subjects.

Continuing the aforementioned work, we proposed a binary classification schema that discriminates among seven emotions, namely: anger, boredom, disgust, anxiety, happiness, sadness, and neutral. The proposed psychologically founded schema has the advantage that it applies the ``divide-and-conquer'' technique, discriminating initially among emotion categories rather than emotions themselves. Also, it is easily adaptable to diverse emotions and its analysis level is adjustable to the problem under consideration. Several features that are gathered, are proposed for the first time within the context of emotion recognition. Feature selection is executed by means of forward feature selection for each gender separately, aiming to take context into account. The Berlin emotional speech database is utilized to conduct speaker-independent experiments. The baseline K-nearest neighbors classifier, as well as support vector machines with linear and Gaussian radial basis function kernel are examined as classifiers. Statistical analysis of the experimental results is carried out with respect to emotion recognition accuracy as well as in terms of confusion matrices evaluation. Analysis verifies that support vector machines are suitable classifiers for emotion recognition under the binary classification schema. In particular, linear support vector machines, which has the advantage of no parametrization, accomplishes a high accuracy of 87.7%, better than the accuracy presented in recent works.

II- Publication(s) during your fellowship

Please insert the title(s), author(s) and abstract(s) of the published paper(s). You may also mention the paper(s) which were prepared during your fellowship period and are under reviewing.

1. M. Kotti, F. Paternò, and C. Kotropoulos, ``Speaker-independent negative emotion recognition", in Proc. *2nd Int. Workshop Cognitive Information Processing,* Elba Island, Italy, June 2010.

Abstract

This work aims to provide a method able to distinguish between negative and non-negative emotions in vocal interaction. A large pool of 1418 features is extracted for that purpose. Several of those features are tested in emotion recognition for the first time. Next, feature selection is applied separately to male and female utterances. In particular, a bidirectional Best First search with backtracking is applied. The first contribution is the demonstration that a significant number of features, first tested here, are retained after feature selection. The selected features are then fed as input to support vector machines with various kernel functions as well as to the K nearest neighbors classifier. The second contribution is in the speaker-independent experiments conducted in order to cope with the limited number of speakers present in the commonly used emotion speech corpora. Speaker-independent systems are known to be more robust and present a better generalization ability than the speaker-dependent ones. Experimental results are reported for the Berlin emotional speech database. The best performing classifier is found to be the support vector machine with the Gaussian radial basis function kernel. Correctly classified utterances are 86.73%±3.95% for male subjects and 91.73%±4.18% for female subjects. The last contribution is in the statistical analysis of the performance of the support vector machine classifier against the K nearest neighbors classifier as well as the statistical analysis of the various support vector machine kernels impact.

2. M. Kotti and F. Paternò, ``Speaker-independent emotion recognition exploiting a large feature set and a binary classification schema", in IEEE Tras. Affective Computing (submitted for publication).

Abstract

This paper deals with the emotion recognition problem from speech signals. A psychologically founded binary classification schema is proposed that divides the problem into successive two-class sub-problems, grouping together emotional categories sharing the same properties. Performance is enhanced since commonly confused in literature pairs of emotions are located away from one another. A large number of features is extracted, related to statistics of pitch, formants, and energy contours, as well as spectrum, cepstrum, perceptual and temporal features, autocorrelation, MPEG-7 descriptors, Fujisaki's model parameters, voice quality, jitter, and shimmer. Among them, several features are tested here for the first time on emotion recognition. To take context into account feature selection is applied for each gender separately. Selected features are fed as input to K nearest neighborhood classifier and to support vector machines. Two kernels are tested for the latter: linear and Gaussian radial basis function. The recently proposed speaker-independent experimental protocol is tested on the Berlin emotional speech database. Speaker-independence is guaranteed by leave-one-speaker-out evaluation, meaning that one speaker cannot appear in both the training and the test set. Speaker-independent systems avoid classifier overfitting and handle efficiently an unknown speaker by preventing speaker adaptation. The best emotion recognition accuracy, achieved by support vector machines with linear kernel, equals 87.7%, outperforming state-of-the-art recent approaches. Statistical analysis is carried out firstly, with respect to the classifiers' error

rates and secondly, to evaluate the information expressed by the classifiers' confusion matrices.

III -Attended Seminars, Workshops, and Conferences

Please identify the name(s), date(s) and place(s) of the events in which you participated during your fellowship period.

1. Summer School: *Cognitive Science and Machine Learning,* Pula, Italy, May 2010 http://www.mlss.cc/sardinia10

2. *2nd Int. Workshop Cognitive Information Processing,* Elba Island, Italy, June 2010 http://www.conference.iet.unipi.it/cip2010/