

Agent-Based Characteristic Assortment using opt-aiNET Algorithm for Sensation Gratitude

K.R.Ananth

Professor, Velalar College of Engineering and Technology, Erode, Taminadu, India

ABSTRACT: Emotion recognition from speech signals has been among the attractive subjects lately. As known the most important process steps were feature extraction and feature selection in Emotion Recognition (ER). In the present study, *Agent-Based Modelling* (ABM) is suggested while selecting the best features for recognition. The proposed method is based on feature selection with optimization algorithm among the features obtained from speech signals. Firstly Mel-Frequency Cepstrum Coefficients (MFCC) were extracted from the emotion database. Several statistical values were obtained from MFCC. Recently Available statistical features were considered as an agent and opt-aiNET optimization algorithm was applied in order to choose the agent group giving the best classification success. Artificial Neural Network (ANN) was used as classifier. As a result of application, it was seen that classification made by using selected features with applied optimization algorithm was more successful after ABM was applied. Proposed method was shown promising performance with MFCC features.

KEYWORDS: Agent based modeling, emotion recognition, MFCC, opt-ai NET

I. INTRODUCTION

From the perspective of human and computer interaction, it can be noticed that Emotion Recognition (ER) has gained importance recently. ER can be made via facial expressions, written text, and biomedical signals or speech signals [1,2]. Especially in the situation when face-to-face communication is not made, determining the emotion state of a person is made from speech data.

Feature extraction and selection are one of the most important steps in ER as in other signals. The feature selection methods can be examined in two ways as spectral and prosodic features. Mel-Frequency Cepstrum Coefficients (MFCC), and Linear Predictive Coefficients (LPC) can be considered as spectral features. Basic frequency (F0) value is one of the main prosodic features. In the present study, data obtained from Berlin Emotion Database for ER were used. Firstly, Mel Coefficients were obtained from the data. Since obtained features had too large dimensions to give them to classifiers and they were in different dimensions, new features were calculated by extracting some statistical features from these data. So as to determine the features which supplies to obtain the high accuracy level, opt-aiNET optimization algorithm on ABM were applied. ABM with opt-aiNET was applied in this study for the first time in literature. Obtained features were given to the classifier. As classifier, Artificial Neural Network was used. Classification results with ABM and without ABM were compared. It was observed that classification accuracy increases using ABM.

The remainder of this paper is structured as follows. In Section 2, information related to used methods was given. In Section 3, database was shortly explained. The experiments to assess the performance of proposed method were described in Section 4. In the last part, conclusion was given.

II. METHOD

Emotion recognition from speech signals has been among the attractive subjects lately. As known the most important process steps were feature extraction and feature selection in emotion recognition (ER). The process of ER can be made from facial expressions and speech signals. Schematic representation of ER process realized in this study

is obtained via the steps depicted in Fig. 1. Although directly ER from signal information seems theoretically possible, the dimension reducing process is needed by using features extracted from data through different methods because data dimension is very large. When it is considered from this point of view, it can be clearly said that the most important steps are feature extraction and selection in emotion detection. Many methods exist for feature extraction in speech processing. The most known and used ones are MFCC, LPC, F0 value, Wavelet transform, Autoregressive Parameters (AR).

2.1. Mel-Frequency Cepstrum Coefficients (MFCC)

Mel-Frequency Cepstrum Coefficients (MFCC) comes first among spectral features that are used most widely to obtain feature from speech signals. MFCC are obtained, the steps of process shown in Figure 2. Input signal is divided into parts in a way that M is sample number and N is sample length ($M < N$). While the first frame consists of N sample, the subsequent frame starts after M sample from the first frame and thus, the samples as $N-M$ match up with. [13] Then, the process of windowing is carried out using the function of Hamming windowing. The function of Hamming windowing is indicated in Formula 1.

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 1 \leq n \leq N \quad \dots(1)$$

The windowed sign is passed through a FIR filter of first degree in pre-emphasis process. Fast Fourier Transform (FFT) is applied in order to transform speech part consisting of N samples from time domain into frequency domain. Mel scale is a scale explained with changing intervals in a way that it changes as linear up to 1 kHz and after 1 kHz, it changes as logarithmic. The logarithm of the sign obtained at the output of Mel filter is taken. Data at the logarithmic Mel scale in Discrete Cosine Transform that is the last step of MFCC extraction process is transformed into time environment again. As a result, data obtained are called as Mel Coefficients. [13]

2.2. Artificial Neural Network (ANN)

ANN is inspired by the work of human neurons. The brain basically learns from experience. ANN can be considered as an electronic model of neural network structure [14]. Neurons that are cells of real neural network in ANN are modelled as a knot. Networks are obtained with binding more than one knot. Multilayer structures consist with binding of two or more neural layers. In real-life problems, multilayer ANN (multi-layer perceptron) is commonly used. ANN is applied widely and efficiently in many pattern recognition problems, speech recognition, and learning robot control strategies [15]. In this paper ANN was applied for classification.

2.3. Opt-aiNET algorithm

aiNET algorithm is a discrete immune network algorithm that is developed for clustering. Opt-aiNET algorithm is a little more developed state of it and its adaptation for optimization of problems [16]. The optimization version of aiNET (Opt-aiNET) can be summarized as follows [16]. Network cell: individual of the population. Fitness: fitness of a cell in relation to an objective function to be optimized. Affinity: Euclidean distance between two cells. (Eq. 2)

$$(XY) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad \dots(2)$$

2.4. Agent-based modelling (ABM)

Although agent does not have an exact definition, it can be defined as an object having features of target, action and state in a particular environment [17]. Moreover, it can be stated as a computer system that can act automatically in order to fulfill a particular aim (Figure 3). An autonomous agent is defined as a system that receives (virtual or real) perception of the environment where it exists, creates situational awareness after that and by using this perception information in accordance with the aim, it determines its subsequent behavior and that realizes determined action in environment[18]. In a similar way, extracted feature groups are considered as agents and optimization process is considered as an environment. Action value of agent is determined according to classification success. If the classification success is maximum, state is being sent as "1" to the agent and the agent is used for ER, with the action value "1". Otherwise, it is decided that *state* will not be classified because *state* is "0" and at the same time, *action* value of the agent is "0".

III. MATERIAL

3.1. Berlin Database of Emotional Speech (Emo-DB)

Many databases are applied in ER problems. It is possible to classify them in different ways according to their number of emotion included, with which language they are formed, as being public or private, their voicing by professionals or actors [19]. A list of the most important database according to their speaking language[20] is also given. The access of databases created by being vocalized by professionals is expensive because they are generally private. LDC Emotional Prosody Speech and Transcripts database language simulated the emotions. [19]. The emotions in this database are Neutral, panic, anxiety, hot anger, cold anger, despair, sadness, elation, Joy, interest, boredom, shame, pride and contempt. INTERFACE[21], is one of the most used databases. Actors stimulated the emotions of Anger, disgust, fear, Joy, surprise, sadness, slow neutral, fast and neutral emotions in English, Slovenian, Spanish, and French languages.

IV. RESULTS AND DISCUSSION

Features obtained by using EMO-DB database in this application were modelled as agent-based and ER was carried out with ANN based on opt-aiNET optimization algorithm. The process steps carried out are indicated in Figure 4. Firstly, statistical values were calculated from Mel Coefficients obtained by using FCC methods on emotion data. In second stage of study, to select the features which increase the classification accuracy ABM was used. As optimization algorithm for feature selection Opt-aiNET algorithm was used. Finally, selected features were used as inputs for ANN and emotion classification was done. In this study, ANN classification of Weka[22] (an open-source public available toolbox for automatic classification) was used. Multilayer Perceptron Algorithm (MLP) was preferred. As a result of experiments, learning rate and momentum coefficient were determined as 0.2 and 0.1, respectively. Seven neurons were used in hidden layer in the application. 10-fold cross validation was applied in order to indicate reliability of the study. In advance, 16 Mel-Coefficients of different lengths from each data were obtained. In the first step, data reduction was made by taking statistics of each coefficient. Extracted statistical values are indicated in Table 2. Matrix dimensions at the steps of MFCC feature extraction process for a randomly selected segment were indicated in Figure 5. As seen in the figure, the segment dimension was 30896 at the first stage and then Mel coefficients were extracted, so the dimension was obtained as [388x16]. In other words, 16 Mel coefficients were obtained. Finally, new fixed-dimensional feature vector of [7x16] by calculating statistical features for each Mel coefficient is given in Table 2.

As a result, a new data set consisting of features of 112 [7x16] for each segment by transforming this matrix into row matrix was obtained. This data set whose dimension is 520 segments with 112 features [520x112] was defined as *Dataset1*. Before feature selection was applied, *Dataset1* was classified with ANN method and classification success was obtained as 72.49%. Confusion matrix obtained after classification of *Dataset1* with ANN is indicated in Table 3. When Table 3 is examined and first line is taken, it is seen that 81 segments in total were taken for the emotion of *Boredom*, 53 of these data were classified as emotion of *Boredom* and the rest of segments (28 segments) were classified as different emotions. This means that emotion of *Boredom* was determined as 65.43% of accuracy. It is

observed that two of 28 segments which cannot be classified accurately was classified as *Disgust*, six of them as *Sadness*, three of them as *Fear*, four of them as *Happiness* and 13 of them as *Neutral*. The emotions that are confused most with emotion of *Boredom* are seen to be *Neutral* and *Sadness* from the Table 3. When other emotions are examined, emotions of *Happiness*, *Anger* and *Fear* are seen to be confused in classification process.

The processes of ABM and optimization were applied in order to observe the effect of feature selection upon classification success. Modelling process is indicated in Figure 6. Each group of statistical values obtained from Mel coefficients was considered as an agent. Maximum value belonging to each group of coefficient was defined as *Agent1*, minimum value as *Agent2*, average value as *Agent3*, Standard Deviation value as *Agent4*, Median value as *Agent5*, Skewness value as *Agent6* and Kurtosis value as *Agent7*. Obtained results and selected agents are shown in Table 8. Classification accuracy of 75.77% was obtained when *Dataset3* was classified with ANN. As for the classification accuracy of *Dataset4*, it was obtained as 78.08%. It was observed that the feature cluster consisting of *Agent3*, *Agent5*, *Agent6* and *Agent7* gave the highest classification success. Confusion matrix obtained as a result of classification is indicated in Table 9.

When Table 9 is examined and *Dataset3* was classified, 104 of *Anger* emotion were accurately classified. When *Dataset4* was used, 106 of segments were classified accurately, which means that the classification success of *Anger* emotion increases from 82.54% to 84.13%. It was observed that the number of defined segments in the emotion of *Happiness* was increased from 43 to 46. Before selection was not made with optimization in the emotion of *Fear*, 50 of them were classified as accurately. By applying the process in Fig. 6, 51 segments were accurately classified.

V. CONCLUSION

Processes made on speech signals are increasing nowadays. Speech recognition, person recognition and ER come first among the studies carried out on speech signals. In this study, the process of ER from emotion data including speech signal was carried out. Features obtained by using Emo-DB database was modelled as agent-based and from the obtained features, the process of ER making feature selection using opt-aiNET optimization was carried out. Mel coefficients were obtained from spectral features and so, statistical values were extracted. The change in classification success after optimization process applied using ABM was analyzed. It was seen that before the approach presented after the tests, the classifying success of 72.49% for the database which includes all of emotions were increased to 75.14%. In the second study carried out dividing emotions into groups of three, it was observed that the success of 75.77% for the first emotion group (AHF) increased to 78.08%.

It was analyzed the success of 74.67% for the second emotion group (BNS) increased to 80.09%. As a result of tests made, it is seen that the presented method gave successful results in feature selection. In this study, the paper aim to try this method on different features.

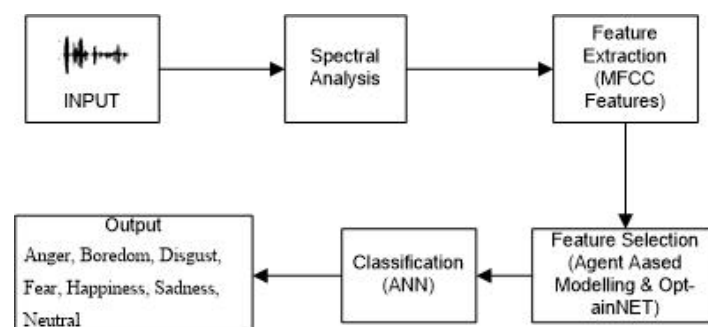


Figure-1: Schematic representation of ER process

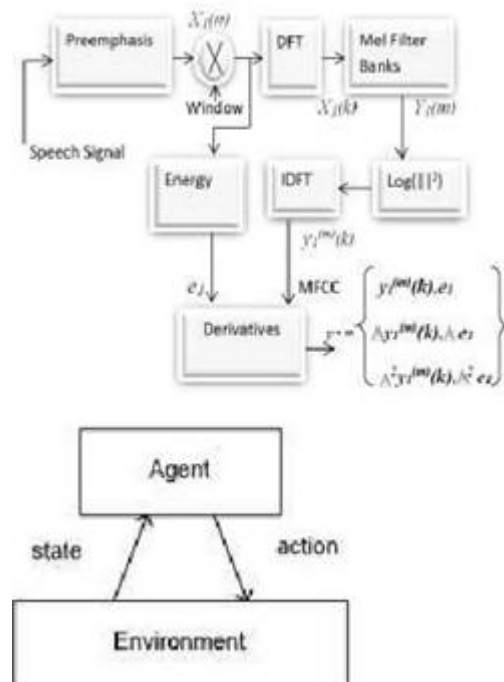


Figure-2: Process steps of MFCC[13]

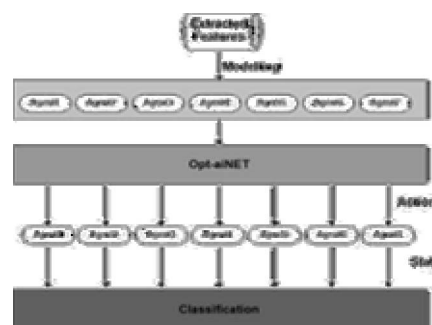


Figure-3: Agent Structure

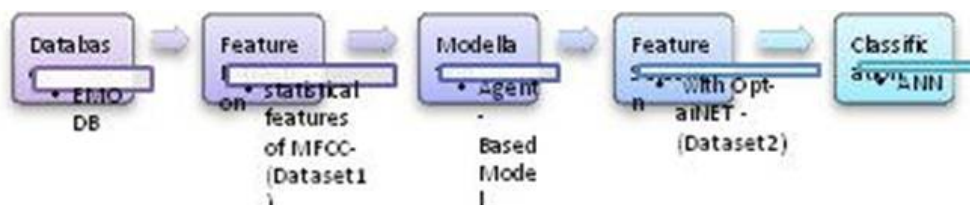


Figure-4: The architecture of speech ER of proposed method

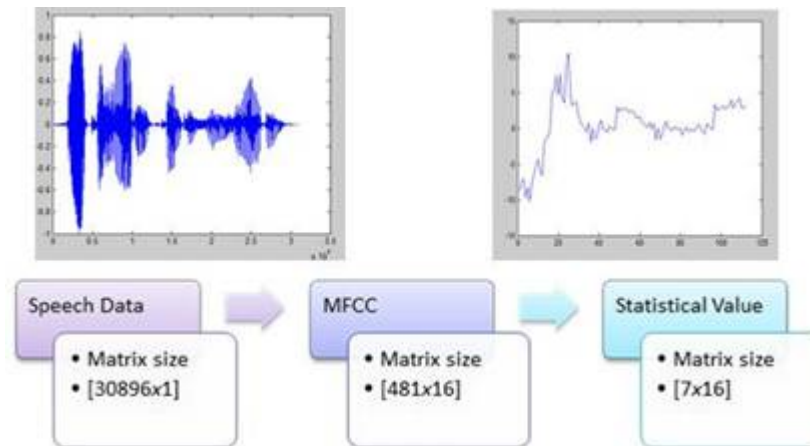


Figure-5: Feature extraction through MFCC by using a randomly selected segment

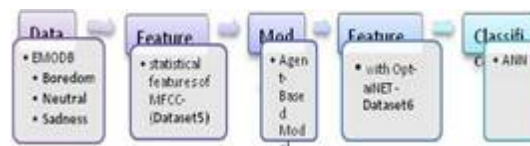


Figure 6: Agent and optimization structure

REFERENCES

- Perikos, I., Hatzilygeroudis, I.: Recognizing emotions in text using ensemble of classifiers.
- Engineering Applications of Artificial Intelligence 51, 191-201 (2016). doi:10.1016/j.engappai.2016.01.012
- Kim, J.-B., Park, J.-S.: Multistage data selection-based unsupervised speaker adaptation for personalized speech emotion recognition.
- Engineering Applications of Artificial Intelligence 52, 126-134 (2016). doi:10.1016/j.engappai.2016.02.018
- Park, J.S., Kim, J.H., Oh, Y.H.: Feature Vector Classification based Speech Emotion Recognition for Service Robots. Ieee T Consum Electr 55(3), 1590-1596 (2009).
- Schuller, B., Rigoll, G., Lang, M.: Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - Belief Network architecture. 2004 Ieee International Conference on Acoustics, Speech, and Signal Processing, Vol I, Proceedings, 577-580 (2004).
- France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., Wilkes, D.M.: Acoustical properties of speech as indicators of depression and suicidal risk. Ieee T Bio-Med Eng 47(7), 829-837 (2000). doi:10.1109/10.846676
- Milton, A., Tamil Selvi, S.: Class-specific multiple classifiers scheme to recognize emotions from speech signals. Computer Speech & Language 28(3), 727-742 (2014). doi:10.1016/j.csl.2013.08.004
- Nwe, T.L., Foo, S.W., De Silva, L.C.: Speech emotion recognition using hidden Markov models. Speech Communication 41(4), 603-623 (2003). doi:10.1016/s0167-6393(03)00099-2
- Albornoz, E.M., Milone, D.H., Rufiner, H.L.: Spoken emotion recognition using hierarchical classifiers. Comput Speech Lang 25(3), 556-570 (2011). doi:10.1016/j.csl.2010.10.001
- Lee, C.C., Mower, E., Busso, C., Lee, S., Narayanan, S.: Emotion recognition using a hierarchical binary decision tree approach. Speech Communication 53(9-10), 1162-1171 (2011). doi:10.1016/j.specom.2011.06.004
- Montano, B.R., Yoon, V., Drummey, K., Liebowitz, J.: Agent learning in the multi-agent contracting system [MACS]. Decision Support Systems 45(1), 140-149 (2008). doi:10.1016/j.dss.2007.12.013
- Taha, M., Helmy, T., Alez, R.A.: Multi-agent Based Arabic Speech Recognition. Paper presented at the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, Marriott Fremont, CA, USA, 02-05 Nov. 2007
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.: A Database of German Emotional Speech. Proc. Interspeech, (2005).
- Becchetti, C., Ricotti, L.P.: Speech Recognition: Theory and C++ Implementation. (2004).
- Anderson, D., McNeill, G.: Artificial Neural Networks Technology In. Kaman Sciences Corporation 258 Genesee Street Utica, New York 13502-4627, (1992)