RESEARCH ARTICLE OPEN ACCESS

Design of an Optimized Emotion Detection System from Speech using Active Feature Selection-Implementation

Suchita V Deshmukh*, Yogadhar Pandey**, ,Arunkumar jhapate**

*(Department of Computer Science & Engineering, SIRT, RGPV, Bhopal, India
Email:-sdsuchitadeshmukh@gmail.com)
**(Assistant Professor ,Deparment of Computer Science & Engineering, SIRT, Bhopal, India
Email:-p_yogadhar@yahoo.co.in)
**(Assistant Professor, Deparment of Computer Science & Engineering, SIRT, Bhopal, India
Email:-arun_jhapate@yahoo.com)

***Abstract*—Data and knowledge management systems employ feature selection algorithms for removing irrelevant, redundant, and noisy information from the data. There are two well-known approaches to feature selection, feature ranking (FR) and feature subset selection (FSS).In this paper, we propose a new FR algorithm, termed as maximum variance selection algorithm, for binary data sets. For data sets having hundreds of thousands of features, feature selection with FR algorithms is simple and computationally efficient but redundant information may not be removed. On the other hand, FSS algorithms analyze the data for redundancies but may become computationally impractical on high-dimensional data sets. We address these problems by combining FR and FSS methods in the form of a two-stage feature selection algorithm. In this paper, For creating featured subset we used activity featured selection (AFS) with maximum variance selection. We apply K-nearest neighbor (KNN) classifier on this featured. Two FSS algorithms are employed in the second stage to test the two-stage feature selection idea. When combined with the FSS algorithms in two-stage it improves their classification accuracy and exhibits up to 97 percent reduction in the feature set size.**

Keywords—Feature ranking, feature subset selection, K nearest neighbor, Active feature selection,.

1.Introduchtion

Emotion recognition from sound has gain a huge popularity in research and development. There is a huge development of personal robots in recent years. Either they may be used for educational purposes or for entertainment purpose. When we see to these robots they may look like familiar pets such as cats or dogs, one such good example is the Sony AIBO robot, or sometimes they may even take shape of the young children's such as humanoids and one good example of such kind of robots is Sony's SDR3-X. The main area of concern here is interaction with these machines. Interaction with these machines are radically different in the way we interact with traditional computers. Human beings learn to use very unnatural conventions and devices such as keyboards or dialog windows and also need to learn computers working in order to use them. On the contrast personal robots have to learn themselves natural conventions such as natural language or social rules such as politeness with appropriate modalities such as speech or touch which human beings are learning from thousands of years and each and every human learns and practices the same from childhood. Among all the above mentioned capabilities, the most basic requirement that a personal robot must possess is the ability to grasp human emotions and in particular must recognize the human emotions as well as to express their own emotions. Importance of emotions are not only limited to human reasoning, nut also extends to centralizing social regulation and also in particular to control dialog flows. Emotional communication is both primitive and efficient enough so that we use it a lot when we interact with pets, in particular when we tame them. This is also certainly

what allows children to bootstrap language learning and it should be inspiring to teach robots natural language. Human beings express emotions in the two main ways viz. the modulation of the facial expressions and the modulation of the intonation of the voice. Researches regarding automated recognition of emotion in facial expression is now very rich and researches dealing with speech modality, both for automated production and recognition by machines has only been active for very few years. By the means of this paper we present the results for research which aims of automatically detecting emotions such as Excitement, Contentment, Depression, Anxious from the given sample of sound. The overall research work consists of two phases viz. Training phase and Evaluation phase as explained in Section II. In the training phase the samples of sound is stored in the database along with the values of its Mel Frequency component (MFCC), Linear predictive code (LPC), Short Term Energy (STE), Zero crossing rate (ZCR) and Energy Entropy (EE) and also the type of the emotion of that particular sound as an input by the user. In the evaluation stage the given sound sample is analyzed and the type of the emotion is determined using rious algorithms such as Active Feature Selection algorithm and K-nearest neighbor algorithm.

1.1 Mel Frequency Cepstral Coefficients (MFCC)

MFCC is most extensively used in speech analysis since past few decades and has also gained popularity in music analysis. MFCC is the means by which spectral information in the sound can be represented. Here the changes within each coefficient across the range of the sound are examined. The process of obtaining MFCC involves analyzing and processing the sound according to the following steps:-

1) Divide the signal into frames.
2) Get the amplitude spectrum of each frame.
3) Take the log of these spectrums.
4) Convert to the Mel scale.
5) Apply the Discrete Cosine Transform (DCT).

The Mel scale is based on human hearing and therefore is a perceptual scale.

1.2 Linear Predective Coding (LPC)

Linear predective coding is one of the most powerful techniques used in speech analysis and with the help of Linear Predective coding, good quality speech can be encoded at a low bit rate and also provides extremely accurate estimation of speech parameters. LPC methods are the most widely used in speech coding, speech synthesis, speech recognition, speaker recognition and verification and for speech storage. The basic idea behind Linear Predective coding is that the current speech sample can be closely approximated as a linear combination of past samples,

$$s(n) = \sum_{k=1}^{p} ak\, s(n-k)$$

for some values of p, a_k 's.
For periodic signals with period N_p , it is obvious that,

$$s(n) \approx s(n - N_p).$$

Basic principles of LPC are as follows:-

1) The time-varying digital filter represents the effects of the glottal pulse shape, the vocal tract IR, and radiation at the lips
2) The system is excited by an impulse train for voiced speech, or a random noise sequence for unvoiced speech
3) This 'all-pole' model is a natural representation for non-nasal voiced speech—but it also works reasonably well for nasals and unvoiced sounds.

1.3 Short Term Energy (STE)

The amplitude of the speech signal varies with time. Generally, the amplitude of unvoiced speech segments is much lower than the amplitude of voiced segments. The energy of the speech signal provides a representation that reflects these amplitude variations. Short-time energy can define as,

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]2$$

The choice of the window determines the nature of the short-time energy representation. In our model, weused Hamming window. The hamming window gives much greater attenuation outside the bandpass than the comparable rectangular window,

(1) h(n) = 0.54 − 0.46cos(2pn /(N −1)) , 0 < n < N −1

h(n) = 0 , otherwise

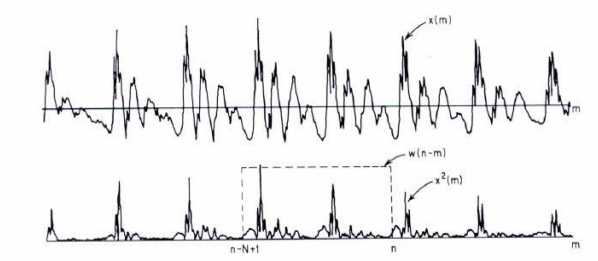


Fig 1: Computation of Short-Time Energy

The attenuation of this window is independent of the window duration. Increasing the length, N, decreases the bandwidth, Fig. If N is too small, E_n will fluctuate very rapidly depending on the exact details of the waveform. If N is too large, E_n will change very slowly and thus will not adequately reflect the changing properties of the speech signal.

1.4 Zero Crossing Rate (ZCR)

In the context of discrete-time signals, a zero crossing is said to occur if successive samples have different algebraic signs. The rate at which zero crossings occur is a simple measure of the frequency content of a signal. Zero-crossing rate is a measure of number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero. Speech signals are broadband signals and interpretation of average zero-crossing rate is therefore much less precise. However, rough estimates of spectral properties can be obtained using a representation based on the short time average zero-crossing rate

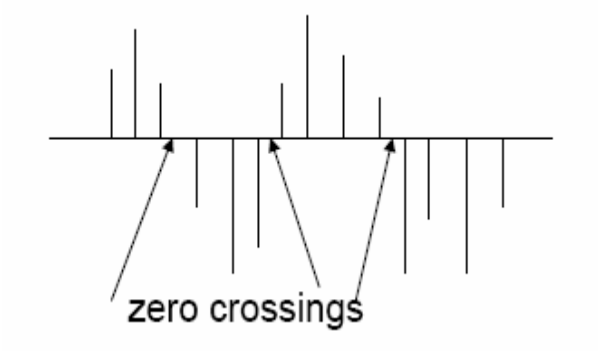


Fig 2: Definition of zero-crossings rate

Zero crossing rate can be defined as,

$$Z_n = |sgn[x(m)] - sgn[x(m-1)]|w(n-m)$$

Where,

$$sgn[x(n)] = \begin{cases} 1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases}$$

And,

$$w(n) = \begin{cases} \frac{1}{2N} \; for, 0 \leq n \leq N-1 \\ 0, \; for \; otherwise \end{cases}$$

The model for speech production suggests that the energy of voiced speech is concentrated below about 3kHz because of the spectrum fall of introduced by the glottal wave, whereas for unvoiced speech, most of the energy is found at higher frequencies. Since high frequencies imply high zero crossing rates, and low frequencies imply low zero-crossing rates, there is a strong correlation between zero-crossing rate and energy distribution with frequency. A reasonable generalization is that if the zero-crossing rate is high, the speech signal is unvoiced, while if the zero-crossing rate is low, the speech signal is voiced.

1.5 K Nearest Neighbor Algorithm

K nearest neighbor algorithm is also called as lazy learning algorithm. This is so because it defers the decision to generalize till a new query is encountered. Whenever we have a new point to classify, we find its K nearest neighbors from the training data.

Algorithm

1)For each training example <x,f(x)>,add the example to the list of training examples.
2) Given a query instance x_q to be classified,
a) Let x1, x2..denote the k instances from training examples that are nearest to x_q.
b)Return the class that represents the maximum of the k instances.

2.PROPOSED WORK

The overall diagrammatic representation of the proposed work is as shown below:-

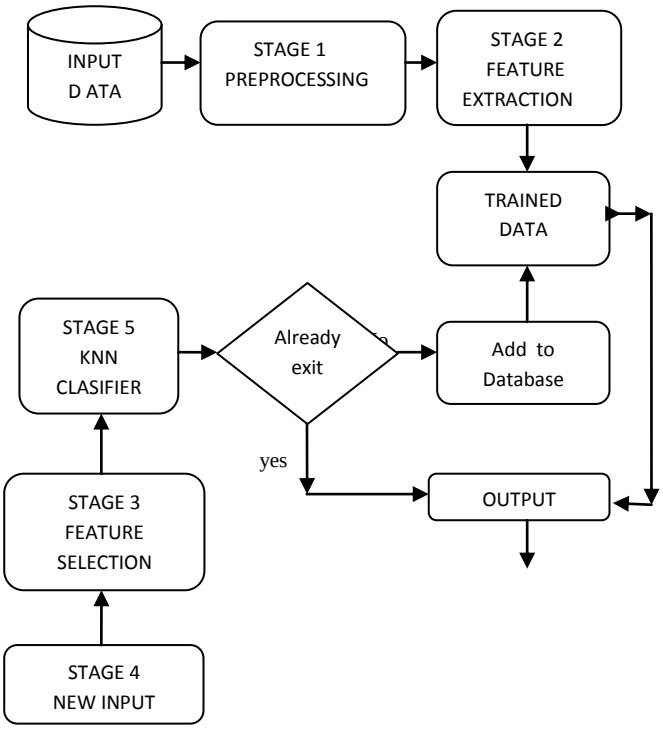


Fig3:- Diagrammatic representation of work flow

The proposed methodology consists of two phases:-

2.1) Training Phase and
2.2) Evaluation Phase

2.1Training Phase

In training phase firstly the complete system is trained. In this phase various sounds as input by the user is saved in the database along with its value of Mel Frequency component or Mel Frequency cepstral coefficients (MFCC), Linear predictive code (LPC), Short Term Energy (STE), Zero crossing rate (ZCR) and Energy Entropy (EE) and also the type of the emotion of each and every particular sound is stored in the database as an input by the user. Either the user can store its own sound or there are some readymade sound databases available. One such database that is readily available is Berkeley database, and the same database is used in our work. The Berkeley is university in US where emotion based research is going on. The second phase consists of an Evaluation phase which is as explained below.

2.2 Evaluation Phase

In the evaluation phase firstly the sound features are given to Active feature selection to get active features from the sound sample. If the active features are more then 2000, then the variance of all the features are calculated and only those features are stored in the database whose value of variance is high. Now the active features are given as input to the system as shown below

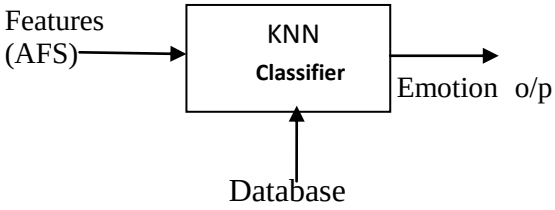
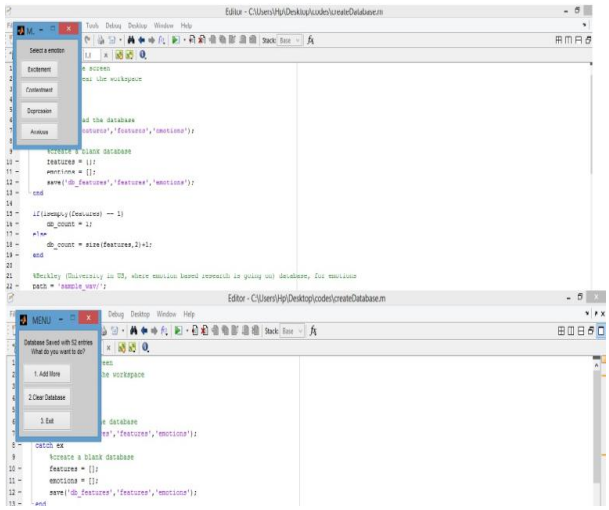


Fig4:-Diagrammatic Representation of Evaluation Phase

Now the active features are given as input to the system. Here KNN algorithm is applied to the input which gives the output as emotion of the particular sound sample. Some of the examples of emotions obtained in output consists of Excitement, Contentment, Depression, Anxious.

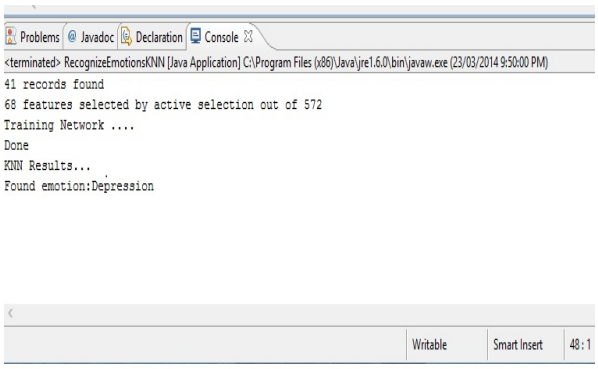
3.MODELING RESULT

STEP 1:- For creating database



This is training phase. As early we define, here we take various input sound by the user and select its type of emotion and saved in database. Advantage of this technique is, size of database if not fixed and it manually created.

Stage 2:- Recognizing new input type of emotion



After creating the database we are ready to get new input type sound to find its emotion. This phase is called as evaluation phase. Where new input sound is check with database calculate its variance value and get output with its emotion type.

4.CONCLUSION

The proposed method brought forward by means of this paper work effectively and efficiently in recognition of emotion from the particular sound sample. Various efficient algorithms are used in the work, thus giving fruitful results. The overall research work consisted of two phases viz. training and evaluation wherein training phase consists of training of the system by user itself feeding the database with particular sound along with its emotion. Hence, the results obtained by this research work is a far more better than the previous works and the proposed method does excellent job in detecting emotion from particular sound sample with high accuracyrate.

REFERENCES

Journal Papers:-

[1]. Guyon and A. Elisseeff, An Introduction to Variable and Feature Selection J. Machine Learning Research, vol. 3, pp. 1157- 1182, 2003.
[2]K. Kira and L.A. Rendell,"A Practical Approach to Feature Selection," Proc. Ninth Int'l Conf. Machine Learning, pp. 249-256, 1992.
[3]A.L. Blumand P.Langley, Selection of Relevant Features and Examples in Machine Learning Artificial Intelligence, ElsevierB.V., vol. 97, pp. 245-271, 1997.
[4]I.C.ASML team,Feature Selection with Redundancy Elimination + GradientBoosted Trees,FactSheet http://clopinet.com isabelle/Projects //agnostic/, 2007.
[5]L. Yu and H. Liu,Efficient Feature Selection via Analysis of Relevance and Redundancy,J. Machine Learning Research, vol. 5, pp. 1205-1224,2004.
[7]M. Dash and H. Liu, "Feature Selection for Classification Intelligent Data Analysis, Elsevier Science B.V., vol. 1, no. 3, pp. 131-156, 1997.
[8] Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh, Feature Extraction, Foundations and Applications. Springer 2006.
[9]M. Hall, "Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning," Proc. 17th Int'l Conf. Machine Learning, 2000.
[10]R. Ruiz and J.S. Aguilar-Ruiz, "Analysis of Feature Rankings for Classification," Proc. Int'l Symp.Intelligent Data Analysis (IDA), pp. 362-372, 2005.
[11]"The Use of Mel-frequency Cepstral Coefficients in Musical Instrument Identification" by Róisín Loughran Jacqueline Walker Michael O'Neill Marion O'Farrell University of Limerick, Limerick, Ireland University of Limerick, Limerick, Ireland University College Dublin, Dublin, Ireland University of Limerick, Limerick, Ireland
[12]Speech Emotion Recognition: Comparison of Speech Segmentation Approaches by Muharram Mansoorizadeh Electrical and Computer Engineering Department Tarbiat Modarres University Tehran, Iran mansoorm@modares.ac.ir Nasrollah.M. Charkari Electrical and Computer Engineering Department Tarbiat Modarres University Tehran, Iran charkari@modares.ac.ir