

An Examination of Emotion Recognition using Machine Learning Algorithms on Different Speech Databases

Kogila Raghu, M Sadanandam, V Kamakshi Prasad

Abstract: *The speech recognition system plays a vital role in understanding the emotions of natural language. The identification of emotions from speech is a challenging task. The performance of the speech recognition system is effects on the speech signals. The speech contains different emotions feelings. Many researchers introduced different emotion recognition techniques. However, these techniques achieved better performance but unsatisfied in identify emotion of natural languages. This paper proposed a novel speech recognition system, which identify the emotions based on the speech signals. The Mel Frequency Cepstral Coefficients (MFCC) features. On the resultant features of speech applied cross-validation using the test emotions.*

The performance of the proposed system verify with the SVM and other two classifiers. The proposed emotion recognition system achieves better performance. The empirical results shows that the proposed system outperforms when compare with different classifiers and databases.

Keywords: Corpora, Features LPCC, MFCC, SVM, HMM, GMM.

I. INTRODUCTION

The information technology is more advanced in emotion recognition of speech. This technology is widely used in many emotion recognition applications. The speech is an essential communication medium in human life. The speech coder is used for emotion recognition of speech. The speech coder process the speech, analysis the speech and recognition speech. The speech coder converts the speech from analog to digital form. Which is better convenient for storage and transmission? The speech coder also converts a digital signal to analog. In the speech, technology implemented many applications and achieved improved performance. However, performance depends on the availability of training resources. Many applications are developed but the performance is limited by the area of resources or language of resources [1].

The emotion recognition applications are developed vastly. The aware of emotion recognition in speech is widely improved. The speech of every human carries different emotions. The performance of an emotion recognition system is affected by extracting relevant information from the speech signals. Generally the emotion recognition had two phases, such as training phase and testing phase.

Revised Manuscript Received on March 18, 2020.

K Raghu, Research Scholar, Department of Computer Science & Engineering, Kakatiya University, Warangal, India.

Dr M Sadanandam, Asst. Professor, Department of Computer Science & Engineering, Kakatiya University, Warangal, India.

Prof. V Kamakshi Prasad, Professor, Department of Computer Science and Engineering, JNTUH College of Engineering Hyderabad, JNT University, Hyderabad, India.

In the first training phase, here different emotions collected information from available resources and store in an emotion recognition database. Second is the testing phase, in which speech is taken to test with trained data. The emotion recognition depends on the features of speech signals. In emotion recognition of speech, the feature extraction mechanism plays a vital role [2].

The MFCC features extraction gives better performance in emotion recognition. Emotion recognition of every speaker is independent. The MFCC features differentiated with different classes. The consonants features are called class level spectral features. The consonants regions of speech have more emotional information compares to stressed or unstressed vowel regions.

Section II discusses previous research works related to proposed work while Section III gives details of research methodology. The experimental results and comparisons discuss in Section IV, the conclusion of research work and future work describe in Section V.

II. RELATED WORK

The performance of emotion recognition in speech depends on the features of speech signals. The different speech applications are tested by the features of linear prediction and cepstral for analysis and classification of emotions in speech.

Xiaoqing Lieut al.[3] In this paper introduced a model for recognition of human emotional state, that is called a novel probabilistic information fusion model. The multi-model emotion recognition framework had rules of data analysis and modules of emotion recognition. In emotion recognition different modules available such as facial expressions, voice analysis, and body posture analysis. The multi-model framework purely works on fusing expression and vice analysis.[3]

Aswin K.M et al.[4] introduced the Human emotion recognition system to recognize the emotion using facial expression and speech signals. In the case of visual expression, the system converts videos into sequence of images the system dynamically detects the emotion of the person. In case of audio speech signals using MFCC, it converts into small frames and extracts the emotions.

A.Revathy et al.[5] introduced Hidden Markov Model Tool Kit for emotion recognizes using MFCC features. In this paper utilized additional noise removal techniques introduced. In the process of pre-processing different noise, techniques are used to remove the noise. Four different emotion recognition model was implemented.

The HMM tool kit achieved better performance in identifying different emotional. Even the proposed tool score good performance but without using the adaptive RLS filter noise removal techniques it degraded.

Yevgeniy Bodyanskiy et al [6] proposed a neo-fuzzy system for online emotion recognition. This system implemented learning algorithm to distribute the emotion data between the different emotion clusters. This system developed based on the entropy criteria and its improved emotion cluster properties also make easy of numerical implementation.

III. PROPOSED METHODOLOGY

Problem Statement:

Emotion Recognition in speech is the scope of many research works. Every research on speech participates in the identification of emotions which is realized to human emotions. The performance of emotion recognition depends on the information collected from the specific emotions and pattern extractions. In real-time speech applications, automatic identifications of emotions from speech signal of a person is necessary. The various classification algorithms and techniques play vital in classifications of these emotional patterns. Many researchers are introduced different algorithms and techniques for emotion to recognize in speech. Those techniques are performed well for international languages and no suitable techniques for linguistic languages.

Process Of Emotion Recognition:

The emotion recognition in speech involves different process. Those are pre-processing, feature extraction, test with trained data and classification. In process of emotion recognition in speech taking database input. The noise is remove from the speech using data pre-processing method. The adaptive pre-processing techniques are used for better noise removal. The pre-processing method gives cleaned speech database. In feature extraction using MFCC Features. The MFCC features works efficiently and accurately. This feature better represents The characteristics of speech signals. Based on the frequency level the main characteristics of speech are extracted. The linear filters at low frequencies and logarithmically at high frequencies. The Hidden Markov Model used for produce trained data. This model gives the maximum successful trained data and produce better accuracy percentage.

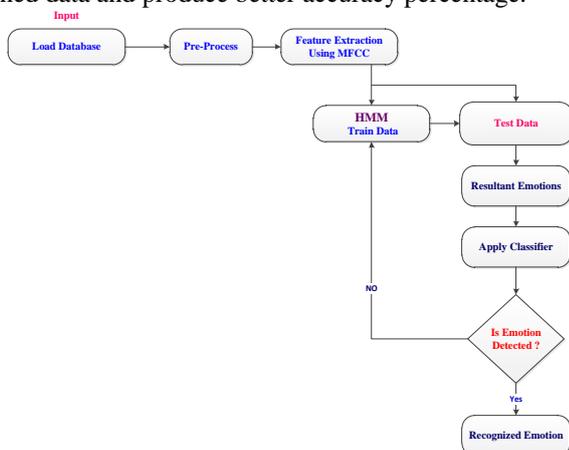


Fig 1: Emotion Recognition in Speech Architecture

The input speech data is tested with trained data after completion of feature extraction by the MFCC feature extraction. The trained data is prepared by using HMM. The feature extracted speech is tested with trained data. Some of the emotions identified in the testing phase. The resultant identified emotions are necessary to classify. Here the classifier plays an important role.

Classification Of Emotions:

The accuracy of emotion recognition mainly depends on the selection of classifiers. The SVM gives better results even in case of limited trained data. The SVM classifier is used to classify the resultant emotions to detect which class of emotions. The proposed system works with different databases using different classifiers.

IV. RESULT ANALYSIS

In this research work proposed a novel emotion recognition system. The proposed system recognizes the different states of emotions. The results of the proposed system are compared with different classifiers on different databases. The implementation of the proposed system and the experiments have been done on a 1.1 GHz Intel Core 2 Duo laptop with 8 GB of RAM. The Python/Matlab programming language is used to implement a novel emotion recognition system. The proposed system achieved high performance. The results of the proposed recognition system are compared to different types of databases using different classifiers. Table I compares the different classifiers performance on the different type of databases.

Table 1: Different Classifiers Performance on Different Data Bases

S No	Classifier Name	Type of Database		
		TESS	RAVDESS	SHEMO
1	k-NN	63.57%	45.44%	49.53%
2	LR	77.71%	47.88%	43.12%
3	SVM	85.14%	52.11%	56.20%

As shown in Table I, the proposed system showed better performance with the TESS database compares to RAVDESS and SHEMO. The SVM classifier is showed better accuracy with comparisons of other classifiers.

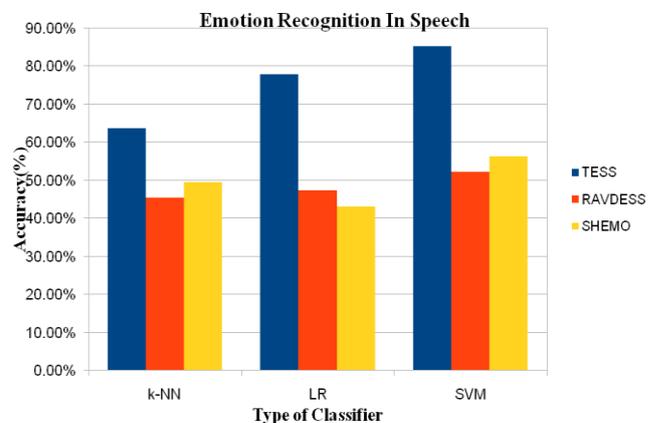


Fig 2 Performance comparison of emotion recognition

The accuracy of emotion recognition using different databases is compared using classifiers. As shown in Fig 2 rate of emotion recognition of the proposed system using different classifiers. The proposed system achieved high accuracy results with the comparison of existing research work.

V. CONCLUSION

In this paper proposed a novel emotion recognition system. However, the demand for emotion recognition from the human speech is increasing, as well as its application area is also expanding. Many researchers are introduced different emotion techniques, but these techniques are limited to languages and sources. In this research work proposed a novel emotion recognition system using different speech databases. The proposed system finds different emotions from input speech data. The MFCC feature extraction used to extract the features from the pre processed speech data. The computed features data are going to the testing phase to identify the emotions. In this research work, much real-time speech data is used for emotion recognition. In this research work, different classifiers are used for the classification of emotions. In the future, this research work is expanded to identified more emotions in the speech signals and improve performance.

REFERENCES

1. P. Vijai Bhaskar et al. (2013). EMOTIONAL TELUGU SPEECH SIGNALS CLASSIFICATION BASED ON K-NN CLASSIFIER. International Journal of Research in Engineering and Technology. 2 (1), p85-93.
2. Naga Padmaja. Jagini et al. (2017). Exploring Emotion Specific Features for Emotion Recognition System using PCA Approach. IEEE. 1 (1), p58-62.
3. Xiaoqing Liu et al., (2011). A Robust Multi-modal Emotion Recognition Framework for Intelligent Tutoring Systems. IEEE. 1 (1), p63-65.
4. Aswin K.M et al., (2016). HERS:Human Emotion Recognition System. IEEE. 1 (1), p176-179.
5. P.Shanmugapriya et al. (2015). Performance Comparison of Speaker and Emotion Recognition. IEEE. 1 (1), p1-6.
6. Yevgeniy Bodyanskiy et al., (2018). The Multidimensional Extended Neo-Fuzzy System and its Fast Learning for Emotions Online Recognition. IEEE. 1 (1), p473-477.
7. Giannakopoulos, T., Pikrakis, A., & Theodoridis, S. (2009). A dimensional approach to emotion recognition of speech from movies. In Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 65–68).
8. Grimm, M., Kroschel, K., Mower, E., & Narayanan, S. (2007). Primitives-based evaluation and estimation of emotions in speech. Speech Communication, 49(10–11), 787–800.
9. Hamidi, M., & Mansoorizadeh, M. (2012). Emotion recognition from Persian speech with neural network. Artificial Intelligence and Applications, 3(5), 107–112.
10. Heni, N., & Hamam, H. (2016). Design of emotional education system mobile games for autistic children. In Proceedings of the 2nd international conference on advanced technologies for signal and image processing (ATSIP).
11. Huahu, X., Jue, G., & Jian, Y. (2010). Application of speech emotion recognition in intelligent household robot. In Proceedings of international conference on artificial intelligence and computational intelligence (Vol. 1, pp. 537–541).
12. James, A. (1994). Is there universal recognition of emotion from facial expression? A review of the crosscultural studies. Psychological Bulletin, 115(1), 102–141.
13. Johnstone, T., Van Reekum, C., Hird, K., Kirsner, K., & Scherer, K. (2005). Affective speech elicited with a computer game. Emotion, 5(4), 513–518.
14. Keshtiari, N., Kuhlmann, M., Eslami, M., & Klann-Delius, G. (2015). Recognizing emotional speech in Persian: A validated database of

- Persian emotional speech (Persian ESD). Behavior Research Methods, 47(1), 275–294.
15. Kort, B., Reilly, R., & Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In Proceedings of the IEEE international conference on advanced learning technologies (ICALT) (pp. 43–46), Washington, DC, USA.
 16. Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.
 17. Lee, C., Mower, E., Busso, C., Lee, S., & Narayanan, S. (2011). Emotion recognition using a hierarchical binary decision tree approach. Speech Communication, 53(9–10), 1162–1171.
 18. Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, J. R. (2007). Neural correlates of processing valence and arousal in affective words. Cerebral Cortex, 17(3), 742–748.
 19. Livingstone, S., Peck, K., & Russo, F. (2012). RAVDESS: The Ryerson audio-visual database of emotional speech and song. In Proceedings of the 22nd annual meeting of the Canadian Society for Brain, Behaviour and Cognitive Science (CSBBCS), ON, Canada.
 20. Mansoorizadeh, M. (2009). Human emotion recognition using facial expression and speech features fusion. PhD thesis, Tarbiat Modares University, Tehran, Iran (in Persian).

AUTHOR PROFILE



K Raghu is a Research Scholar in the Department of Computer Science & Engineering, Kakatiya University, Warangal, India. He is currently working as an Asst. Professor in CSE, Vignan's Institute of Technology & Aeronautical Engineering, Hyderabad. He did his M.Tech from CBIT, Hyderabad in 2013. His research includes Machine Learning, Pattern Recognition and Data Mining.



and Data Mining.

Dr M Sadanandam working as an Asst. Professor in the Department of Computer Science & Engineering, Kakatiya University, Warangal, India. He received his Ph.D from JNTU, Hyderabad Under the Guidance of Prof. Kamakshi Prasad. He Published more than 30 papers in reputed Journals. He is guiding 4 Scholars in his research discipline. His research includes Pattern Recognition, Computer Vision



Prof. V Kamakshi Prasad is currently working as a professor in the Department of Computer Science and Engineering, JNTUH College of Engineering Hyderabad, JNT University, Hyderabad, India. He received PhD degree in Speech Processing from Indian Institute of Technology Madras (IITM) in 2002, Madras, India. He has about 26 years of teaching experience in academics. He has guided 22 PhD Scholars and guiding 6 more PhD scholars in his research disciplines. He has published about 120 papers in international journals and conferences. His research interests include Speech processing and Recognition, Image processing, Data Mining, Machine Learning and wireless communication networks.