Stochastic Modelling and Computational Sciences

A COMPARATIVE STUDY ON BEHAVIOURAL ANALYSIS OF EMOTION THROUGH SPEECH EMOTION RECOGNITION OF HINDI LANGUAGE

Ms. Sujata Kotian
Ramniranjan Jhunjhunwala College of Arts, Science and Commerce
sujatakotian@rjcollege.edu.in

ABSTRACT

Emotion plays a significant role in daily interpersonal human interactions. This is essential to our rational as well as intelligent decisions. It helps us to match and understand the feelings of others by conveying our feelings and giving feedback to others. Research has revealed the powerful role that emotion plays in shaping human social interaction. Emotional displays convey considerable information about the mental state of an individual. In prior studies, several modalities have been explored to recognize the emotional states such as facial expressions, speech, physiological signals, etc. SER (Speech Emotion Recognition) aims to recognize the underlying emotional state of a speaker from her voice. The area has received increasing research interest all through current years. There are many applications of detecting the emotion of the persons like in the interface with robots, audio surveillance, web-based E-learning, commercial applications, clinical studies, entertainment, banking, call centres, cardboard systems, computer games, etc. For classroom orchestration or E-learning, information about the emotional state of students can provide focus on the enhancement of teaching quality. For example, a teacher can use SER to decide what subjects can be taught and must be able to develop strategies for managing emotions within the learning environment. That is why the learner's emotional state should be considered in the classroom. Three key issues need to be addressed for a successful SER system, namely, choice of a good emotional speech. Three key issues need to be addressed for a successful SER system, namely, choice of a good emotional speech database, extracting effective features, and designing reliable classifiers using machine learning algorithms. In fact, the emotional feature extraction is a main issue in the SER system. It involves classifying the raw data in the form of utterance or frame of the utterance into a particular class of emotion on the basis of features extracted from the data.

Keywords: SER, Limitations, Emotion Recognition

1. INTRODUCTION

Human computer interaction (HCI) is getting considerable attention from lots of researchers due to its practical applications in ubiquitous systems (Hassan et al., 2019; Yang et al., 2020; Pace et al., 2019; Gravina and Fortino, 2016; Zhang et al., 2018). For instance, adopting HCI systems in a ubiquitous healthcare system can improve it by perceiving people’s accurate emotions and proactively acting to help them improve their lifestyle. Alongside other data sources, research on emotion recognition from audio is increasing day by day for healthcare in a smartly controlled environment. Speech is a natural way for humans to communicate with each other in daily life. In effective computing research, speech has a vital role in promoting harmonious HCI systems and emotion recognition from speech is the first step. However, due to the lack of an exact definition of emotion, robust emotion recognition from audio speech seems to be quite complex. Hence, it demands a lot of research to solve the challenging problems beneath the audio-based emotion recognition (Sonmez and Varol, 2019).

Speech signals carry feelings and intentions of the speaker (Zhao et al., 2018). Speech signal analysis can be done in both time and frequency domains to obtain features to model underlying events (e.g., speaker, meaning of the speech, and emotion recognition) in the signals. Hence, original speech signal and corresponding spectrum diagram can be explored for robust emotion recognition for both the domains. In Trigeorgis et al. (2016), the speech signal in the time domain was used as input and combined with a machine learning model for emotion recognition. In Sivanagaraja et al. (2017), the authors simultaneously applied original speech, multiscale, and multi-frequency signals to predict different emotions. In audio speech signals, the waveform characteristics vary irregularly. While studying the effective identification of speech information, the typical approach is to first use the raw audio signal processing and then followed by learning the extracted features with some machine learning models, for comprehensive pattern recognition or event prediction. Spectrogram analysis of the speech signal is also very common for speech pattern recognition. In that case, the speech signal is windowed to small chunks and then divided into narrowband and broadband spectrum (Loweimi, 2016). Emotion recognition from speech signals based on spectrum may contribute much in the feature engineering process.

2. OBJECTIVES OF STUDY

1. To survey on the existing Speech Emotion Recognition (SER) System for different forms of emotions in Hindi Language.
2. To review the existing research on SER for Hindi language.

3. Study Background (Literature Review) and Significance of Study

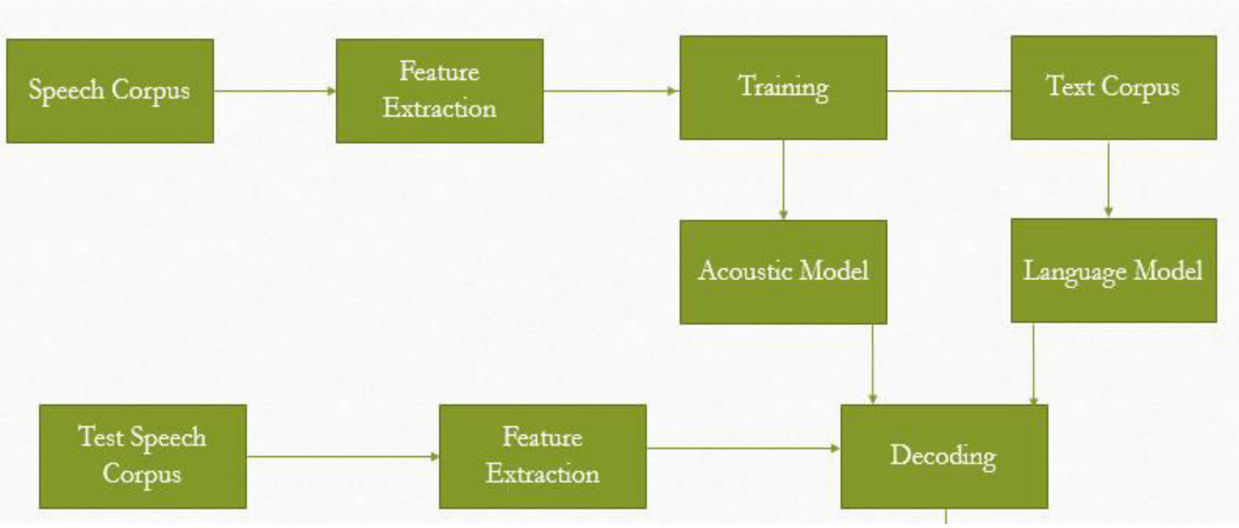
The literature review you provided discusses various research papers related to speech emotion recognition (SER) using machine learning and deep learning techniques. Here's a summary of the key points from these papers and the significance of the study in the field of SER:

- "Applying Machine Learning Techniques for Speech Emotion Recognition": This paper explores the use of Deep Neural Network (DNN) and k-nearest neighbor (k-NN) algorithms for recognizing emotions from speech, particularly scary emotions. It highlights the application of artificial intelligence in this domain.
- "Emotion Recognition using Speech and Neural Structured Learning": The proposed approach in this paper has practical applications, such as understanding emotions in everyday life and in air traffic management systems. It uses Mel Frequency Cepstrum Coefficients (MFCC) and Neural Structured Learning (NSL).
- "Deep Learning Techniques for Speech Emotion Recognition": This paper reviews deep learning and conventional machine learning approaches for SER, considering various techniques like attention mechanisms, autoencoders, CNNs, LSTM, GANs, and more.
- "Machine Learning Based Speech Emotions Recognition System": It focuses on improving emotion recognition systems by adding a deep neural network and utilizes classifiers like K-Nearest Neighbors, Support Vector Machine, Naive Bayes, and Convolutional Neural network.
- "Emotion-based Classification of Human Voice": This paper uses an ensemble of naive Bayes classifiers for binary classification of emotions based on sound feature extraction and machine learning.
- "Extraction of Emotions from Speech - A Survey": The paper reviews popular datasets and classifiers used for automatic emotion recognition in speech, providing insights into the availability and size of datasets.
- "Speech Emotion Recognition using Deep Learning": It addresses issues related to databases and methodologies for emotion recognition and utilizes Inception Net with IEMOCAP datasets for emotion recognition.
- "Speech-based Emotion Recognition using Machine Learning": This paper focuses on recognizing emotions in speech and classifying them into six emotion classes using classifiers like Support Vector Machine, Random Forest, and Convolutional Neural Network.
- "Speech Emotion Recognition using Deep Learning Techniques: A Review": This review covers various deep learning techniques applied to SER, emphasizing contributions, limitations, and database usage.
- "Learning Salient Features for Speech Emotion Recognition": The paper proposes a method to learn affect-salient features for SER using convolutional neural networks (CNNs) and other techniques.
- "Techniques and Applications of Emotion Recognition in Speech": It provides an overview of techniques such as Artificial Neural Networks, k-NN, Support Vector Machines, and probabilistic models for emotion recognition in speech.

- "Speech Emotion Recognition Using Speech Feature and Word Embedding": This paper combines text and speech features to improve emotion recognition accuracy.
- "An Improved Hindi Speech Emotion Recognition System": It explores emotion recognition in Hindi using statistical and neural network techniques.
- "EmoInHindi: A Multi-label Emotion and Intensity Annotated Dataset in Hindi": The paper introduces a dataset for multi-label emotion and intensity recognition in Hindi conversations.
- "Identification of Hindi Dialects and Emotions using Spectral and Prosodic Features of Speech": This research deals with identifying Hindi dialects and recognizing emotions using speech features.
- "Speech Emotion Recognition of Hindi Speech using Statistical and Machine Learning Techniques": The paper combines different types of speech features and machine learning techniques for emotion recognition in Hindi speech.
- "Acoustic Analysis and Perception of Emotions in Hindi Speech using Words and Sentences": This study investigates perceptual evaluation of emotions in Hindi speech and their acoustic correlates.
- "Emotions in Hindi Speech - Analysis, Perception, and Recognition": The paper explores the analysis, perception, and recognition of emotions in Hindi speech, highlighting the importance of acoustic parameters.
- "Feature Extraction Techniques with Analysis of Confusing Words for Speech Recognition in the Hindi Language": This research focuses on building a speaker-independent connected word Hindi speech recognition system and conducts comparative analysis of confusing words.

The significance of these studies lies in advancing the field of speech emotion recognition, which has numerous applications, including mental health monitoring, customer service, human-computer interaction, and more. These papers contribute by introducing novel techniques, datasets, and insights, helping to improve the accuracy and applicability of SER systems across different languages and domains.

4. PROPOSED METHODOLOGY



- The raw acoustic voice is first converted to signal form, from the signal the feature extraction is made. The features extracted are first pre-processed to fit into the given constraints, then the pre-processed data is trained through Acoustic Model (HMM) which feature is selection model training, classifier is done and to predict next given sequence a probabilistic base language model is used. After identifying different utterances, fruitful output over the precise emotion is obtained.

- Understanding the mood of the person in a direct conversation is just an identification, whereas the detection of mood in an indirect conversation is intelligence. For this intelligence machines require some parameters Such as frequency, pulse, amplitude, structure, harmonic, pitch.

✔ **Frequency:** Variation in the pitch of voice

✔ **Pulse:** Standard deviation in voice that indicates the rate of speaker

✔ **Amplitude:** Variation in loudness of voice

✔ **Structure:** Convey the voiced or unvoiced frame structure.

✔ **Harmonic:** relative highness or lowness of voice

✔ **Pitch:** conveys the mean of the voice and peaks of the sound spectrum of voice.

5. LIMITATIONS OF THE STUDY

1. Limitations of deep learning techniques include their large layer-wise internal architecture, less efficiency for temporally-varying input data and over-learning during memorization of layer-wise information.
2. The positive aspect of CNNs is to learn features from high-dimensional input data, but on the other hand, it also learns features from small variations and distortion occurrences and hence, requires large storage capability.

6. BIBLIOGRAPHY

1. Hassan, M. M., Alam, M. G. R., Uddin, M. Z., Huda, S., Almogren, A., & Fortino, G. (2019). Human emotion recognition using deep belief network architecture. Information Fusion, 51, 10-18.
2. Yang, J., Wang, R., Guan, X., Hassan, M. M., Almogren, A., & Alsanad, A. (2020). AI-enabled emotion-aware robot: The fusion of smart clothing, edge clouds and robotics. Future Generation Computer Systems, 102, 701-709.
3. Pace, P., Aloi, G., Gravina, R., Caliciuri, G., Fortino, G., & Liotta, A. (2018). An edge-based architecture to support efficient applications for healthcare industry 4.0. IEEE Transactions on Industrial Informatics, 15(1), 481-489.
4. Pace, P., Aloi, G., Gravina, R., Caliciuri, G., Fortino, G., & Liotta, A. (2018). An edge-based architecture to support efficient applications for healthcare industry 4.0. IEEE Transactions on Industrial Informatics, 15(1), 481-489.
5. Zhuang, X., Huang, J., Potamianos, G., & Hasegawa-Johnson, M. (2009, April). Acoustic fall detection using Gaussian mixture models and GMM supervectors. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 69-72). IEEE.
6. Sonmez, Y. U., & Varol, A. (2019, June). Legal and Technical Aspects of Web Forensics. In 2019 7th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-7). IEEE.
7. Sivanagaraja, T., Ho, M. K., Khong, A. W., & Wang, Y. (2017, December). End-to-end speech emotion recognition using multi-scale convolution networks. In 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 189-192). IEEE.
8. Loweimi, E., Barker, J., & Hain, T. (2016, September). Use of generalised nonlinearity in vector taylor series noise compensation for robust speech recognition. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH (Vol. 2016, pp. 3798-3802). Sheffield.
9. Subham Banga, Ujjwal Upadhyay, Piyush Agarwal, Aniket Sharma and Prerana Mukherjee, 2019.
10. Banga, S., Upadhyay, U., Agarwal, P., Sharma, A., & Mukherjee, P. (2019). Indian EmoSpeech Command

Dataset: A dataset for emotion based speech recognition in the wild. arXiv preprint arXiv:1910.13801.

11. Banga, S., Upadhyay, U., Agarwal, P., Sharma, A., & Mukherjee, P. (2019). Indian EmoSpeech Command Dataset: A dataset for emotion based speech recognition in the wild. arXiv preprint arXiv:1910.13801.
12. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005, September). A database of German emotional speech. In Interspeech (Vol. 5, pp. 1517-1520).
13. Steidl, S. (2009). Automatic classification of emotion related user states in spontaneous children's speech (p. 250). Berlin, Germany: Logos-Verlag.
14. Batliner, A., Steidl, S., & Nöth, E. (2008). Releasing a thoroughly annotated and processed spontaneous emotional database: the FAU Aibo Emotion Corpus.
15. Schuller, B., Steidl, S., Batliner, A., Schiel, F. and Krajewski, J. "INTERSPEECH 2011 speaker state challenge. In: Interspeech'11," Proc. Interspeech11, 3201-3204 (2011).
16. Mao, Q., Dong, M., Huang, Z., & Zhan, Y. (2014). Learning salient features for speech emotion recognition using convolutional neural networks. IEEE transactions on multimedia, 16(8), 2203-2213.
17. Lugović, S., Dunđer, I., & Horvat, M. (2016, May). Techniques and applications of emotion recognition in speech. In 2016 39th international convention on information and communication technology, electronics and microelectronics (mipro) (pp. 1278-1283). IEEE.
18. Atmaja, B. T., Shirai, K., & Akagi, M. (2019, November). Speech emotion recognition using speech feature and word embedding. In 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 519-523). IEEE.
19. Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A review. IEEE Access, 7, 117327-117345.
20. Deshmukh, G., Gaonkar, A., Golwalkar, G., & Kulkarni, S. (2019, March). Speech based emotion recognition using machine learning. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 812-817). IEEE.
21. Reddy, A. P., & Vijayarajan, V. (2017). Extraction of emotions from speech-a survey. International Journal of Applied Engineering Research, 12(16), 5760-5767.
22. Behera, R. N., Baral, P., Saha, S., & Dash, S. Emotion based Classification of Human Voice using an Optimized Machine Learning Approach.
23. Kumar, Y., & Mahajan, M. (2019). Machine learning based speech emotions recognition system. International Journal of Scientific and Technology Research, 8(7), 722-729.
24. Abbaschian, B. J., Sierra-Sosa, D., & Elmaghraby, A. (2021). Deep learning techniques for speech emotion recognition, from databases to models. Sensors, 21(4), 1249.
25. Uddin, M. Z., & Nilsson, E. G. (2020). Emotion recognition using speech and neural structured learning to facilitate edge intelligence. Engineering Applications of Artificial Intelligence, 94, 103775.
26. Tarunika, K., Pradeeba, R. B., & Aruna, P. (2018, July). Applying machine learning techniques for speech emotion recognition. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.
27. Shashank, B., Shankar, B., Chandresh, L., & Jayashree, R. (2021). Emotion Recognition in Hindi Speech Using CNN-LSTM Model. In Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough (pp. 13-22). Springer, Cham.

28. Koolagudi, S. G., Reddy, R., Yadav, J., & Rao, K. S. (2011, February). IITKGP-SEHSC: Hindi speech corpus for emotion analysis. In 2011 International conference on devices and communications (ICDeCom) (pp. 1-5). IEEE.

29. Agrawal, S. S. (2011, October). Emotions in Hindi speech-analysis, perception and recognition. In 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA) (pp. 7-13). IEEE.

30. Vikram Singh, G., Priya, P., Firdaus, M., Ekbal, A., & Bhattacharyya, P. (2022). EmoInHindi: A Multi-label Emotion and Intensity Annotated Dataset in Hindi for Emotion Recognition in Dialogues. arXiv e-prints, arXiv-2205.

31. Agrawal, A., & Jain, A. (2020). Speech emotion recognition of Hindi speech using statistical and machine learning techniques. Journal of Interdisciplinary Mathematics, 23(1), 311-319.

32. Bansal, S., Agrawal, S. S., & Kumar, A. (2019). Acoustic analysis and perception of emotions in hindi speech using words and sentences. International Journal of Information Technology, 11(4), 807-812.

33. Rao, K. S., & Koolagudi, S. G. (2011). Identification of Hindi dialects and emotions using spectral and prosodic features of speech. IJSCI: International Journal of Systemics, Cybernetics and Informatics, 9(4), 24-33.

34. Bhatt, S., Jain, A., & Dev, A. (2021). Feature extraction techniques with analysis of confusing words for speech recognition in the Hindi language. Wireless Personal Communications, 118(4), 3303-3333.