iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.58933

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Feelflow: Delving into Emotional Depths with Generative AI

Er. Ankita Sharma[1], Shafaque Jabeen[2], Piyush Rawat[3], Saloni Sharma[4]
Dept. of CSE Chandigarh University

Abstract: Significant progress has been made in sentiment analysis in recent decades, with a focus on textual data. Nonetheless, the scientific community has not done a great deal to investigate sentiment analysis in the context of audio. By applying a novel method of sentiment analysis to voice transcripts and concentrating on the subtle interpretation of emotions sent by distinct speakers during conversations, this study seeks to close this gap in knowledge. The main goal of this suggested research paper is to create an advanced sentiment analysis system that can communicate with several users in a seamless manner while identifying and assessing the emotional content that each user is conveying through their audio inputs. Advanced approaches like Recurrent Neural Networks (RNN), Long Short Term Memory(LSTM), Teacher Forcing , Encoder- Decoder Model, Tokenization, Gated Recurrent Units (GRU),(Bidirectional Encoder Representations from Transformers),Gradient Boosting Machine.

I. INTRODUCTION

Sentiment analysis is a useful tool for determining how the general public feels about certain issues, what's their mental state and how different person interpret and deal with that. This method is widely
used by websites, applications, and organizations for a variety of reasons like providing them recommendation about something they need, something that can help them. Our goal is to use our current knowledge to create an assistant that is skilled at understanding and absorbing human mental processes during meaningful discussions. Through the extraction of significant phrases and patterns from conversations, this intelligent tool is able to identify the moods and feelings of individuals. By using a variety of analytical methods and extrapolating knowledge from previous information, it assesses the speaker's pitch as well as the speech's content through an integrated sentiment analysis.

The ability to comprehend the emotions and ideas of others has various applications. A key element of this is the development of technology that can recognize and respond to an individual's emotional state. Imagine a device that could sense its owner's mood and adjust its settings to meet their needs and preferences. These technological advancements have the power to significantly enhance the general satisfaction and experience of customers.

Academic institutions are making great efforts to improve the quality of both the text transcriptions and the audio recordings. This project incorporates data from numerous datasets, including those from earlier models, published scientific articles, Twitter comments, and many more. Their goal is to improve through the use of this technology.

In order to understand how consumers interact with the assistance model, our team of researchers has reviewed papers on sentiment analysis. It's challenging to interpret voice of multiple users discussions and determine each one's unique tone. This paper provides a model that uses Ravdess database which includes voice of 60 actors that can readily understand and identify speakers, based on different algorithm various process have been done, at the end satisfying result have been generated. This model looks perfect for users and the associated emotions because their sentiment is compared with actors voice present in ravdess database, based on their pitch by implementing threshold concept, at last graph is generated for pitch variation.

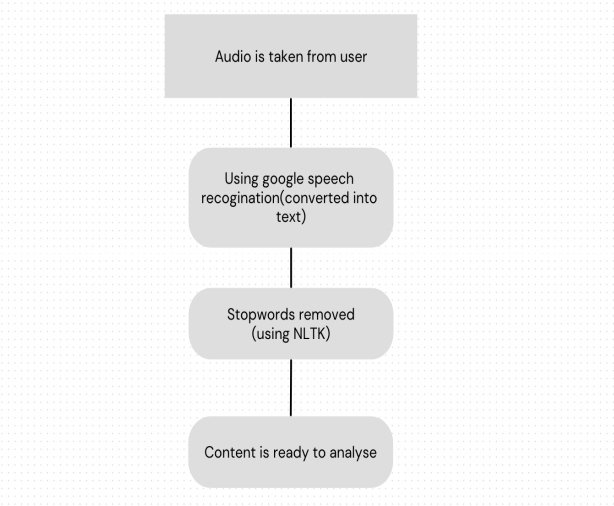
We present an approach named as Recurrent Neural Network, which includes multiple layers such as input layer which is responsible for taking audio input from the user, hidden layer which helps in processing of audio and text to create patterns coupled with suggestions that help to make multiple layers based on current layer and compiling all the data captured after analyzing each and every layer to get a particular sentiment, at last output layer which directly give response to user.

In Part II, we investigate how input is taken from users, removal of extra content, supporting method , Section III provides more detail about tokenization, how processing is done, various supporting methods. While Section IV describes the features of the model. After that, Section V presents the findings and offers a thorough analysis. Episode VI brings the project to a close with confusion matrix.

II. LITERATURE REVIEW AND CONTEXT

A. Taking Input

A microphone is used to record audio. To do this, a number of libraries are required, including librosa, soundevice, and pyaudio, which provide efficient sound playback and recording. To train it on an individual basis, a variety of machine algorithms are utilized, including supervised and unsupervised learning. To maximize accuracy and performance, different ratio training and testing are conducted. We use pre-trained models that are already in existence to extract significant features. Text is created by converting audio to text, which is then ready for analysis through a number of steps.

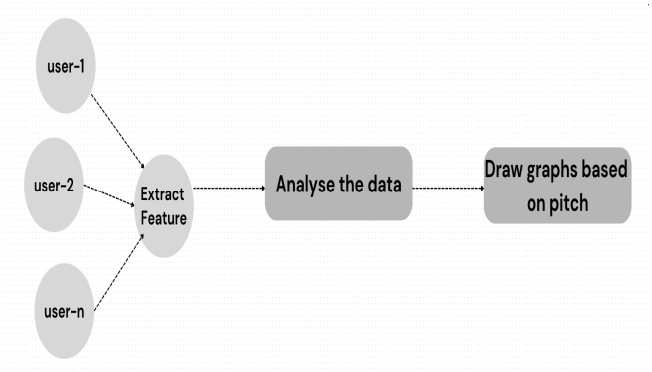


B. Speech Recognition

Google Speech Recognition is used to translate user audio into text. The backend calls a variety of threads to ensure that the operating system is managed in an orderly fashion. For instance, one thread may be used to capture audio, while another is used for conversion, stop word removal, compatibility, or other purposes. Unwanted data can be recorded that doesn't aid in obtaining any kind of crucial information. With the aid of the NLTK library, such content can be removed. The remaining content is finally moved on to the next stage.

C. Initial Processing

It now compares the generated data with our dataset to see if it already exists and responds appropriately; if not, it begins generating its own response. Responses are generated based on analysis and patterns using different machine learning techniques. Pitch information of the audio are retrieved and compared with the actors' voices.Subsequently, the generated data undergoes a series of laborious processes to ensure the generation of very precise and error-free replies. These processes require intricate layers of analysis, refinement, and validation. The goal is to create content that satisfies viewer expectations while showcasing flawless, poignant interactions. The produced data is then put through a number of tedious procedures to guarantee that extremely accurate and error-free responses are produced. Complex layers of analysis, refining, and validation are needed for these procedures. The aim is to produce content that meets the expectations of the audience and has beautiful, touching interactions.



This work develops a speech recognition system using the various algorithms variety of datasets records, pre-trained networks are combined to generate a better response model.

1) Feature Extraction:

This work uses the Ravdess database to generate a new kind of speaker setup. It contains a range of emotions, such as joy, happiness, sadness, calmness, anger, surprise, etc. The user's input is compared with the dataset based on their emotions. Tokens are created from the data on the text side. Every word denotes a token. These tokens are now used to predict words, which aids in the creation of answers.

LSTM feature is used in teacher forcing where you temporary store a token so that it can be used for re-checking as used in teacher forcing to make sure token next predicted word should be perfect. This is a temporary memory so that no extra memory consumption is there, and this block of memory is freed after specific task is done. Teacher forcing is used to double check the things so that we can ensure predicted or generated word is correct according to our requirement.

2) Feature Matching

Image Processing--- Based on the pitch of the input voice graph is created with color variation lying from range of high pitch to the low pitch. Now the created graph is compared with already existed actors expressions graph. If perfect match is found, then from here we are able to detect sentiments of our user. After getting sentiments we fetch data from our dataset. If it is not present, we try to generate it by various techniques. Encoder - Decoder model help to generate response by analyzing the data and then generate it according to user demand. To make sure and verify the created response, each generated response is put through a teacher forcing procedure.

III. PROPOSED SYSTEM

In our research, we often suggest a sentiment analysis model that makes use of decisions made from the provided speech pitch to determine the user's sentiments state. There are eight steps in this approach: The method consists of the eight following steps: preprocessing, feature extraction, model training, language model integration, decoding, post-processing, evaluation, and deployment.

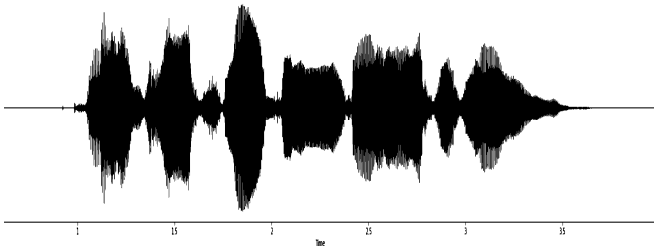
In order to distinguish between received signals and phonetic speech, the system uses trapped signals to identify vocal activity. These noises are handled as discrete pieces of data, which are then sent into the speech recognition and talker discriminating system. This enables people to view the talk's content as well as the speaker. Voice recognition technology then transcribes these segments into text, and systems by speaker ID of the transcripts for additional analysis are compared. Speaker recognition technology uses speaker ID to identify and classify portions, and independently identify them as coming from different loudspeakers from the same loudspeaker. By understanding users' emotions, we may provide alternative terms to make the user experience more comfortable. After utilizing our model, the consumer also feels extremely calm.

We can broaden its appeal and boost its popularity among younger people by utilizing it extensively.

IV. EXPERIMENTAL CONFIGURATION

A. Data set

Within this section of the RAVDESS, there are 1440 files: 24 actors x 60 trials apiece = 1440. With a neutral North American accent, 24 professional actors (12 females and 12 males) perform two lexically matched statements in the RAVDESS. Expressions of peace, happiness, sadness, anger, fear, surprise, and disgust are examples of spoken emotions. Every expression has two emotional intensity levels (strong and normal), in addition to a neutral expression. All 1440 of the files have different filenames. A seven-part number identification makes up the filename (03-01-06-01-02-01-12.wav, for example).



B. Tests and Measurements

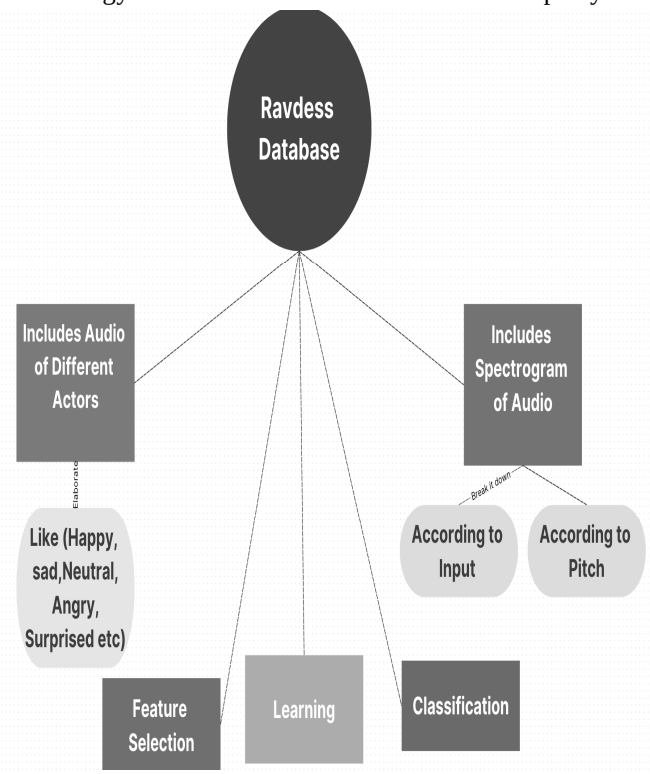
We propose an approach that combines audio analysis, audio recognition, and sentiment analysis. We have provided a helpful analysis of the tests that have been run using different techniques and algorithms. Google Speech API is a voice recognition application programming interface. Many different techniques were used, including LSTM, neural networks, encoder-decoder models, threshold, teacher forcing, and image processing. Additionally, other steps have been done to confirm the results. The system's accuracy has been evaluated using standard sentiment analysis datasets, such as actor voice, speech analysis, and Ravdess, during the sentiment analysis process.

V. RESULTS

A. Results of Audio Recognition

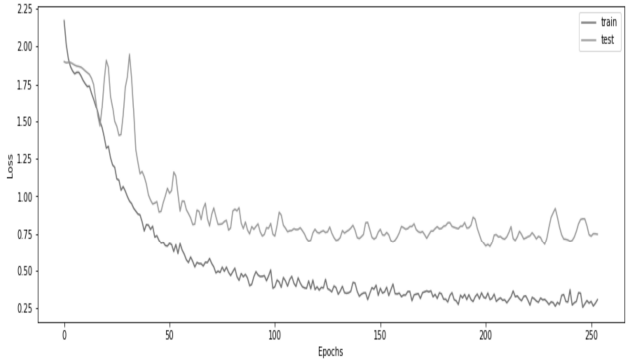
The results of data analysis employing datasets are really fulfilling. If an emotion is recognized as being happy, we can add words like "glad to hear that" or "I'm happy for you" before the response. If an emotion is recognized as being sad, we can add words like "it's okay dear" or "you will definitely get it next time" before the response.

By including these components, users will experience a friendly environment where the system truly understands their needs and feelings. Increasing the size of datasets and improving the user interface ensure accurate results and promote platform flexibility. With the help of technology, consumers can interact with devices and apps more naturally and intuitively, enjoying a seamless, customized experience while controlling devices and making calls, among other tasks. This user-centric approach aims to create a connection between the user and the technology to enhance overall satisfaction and empathy and understanding in interactions.



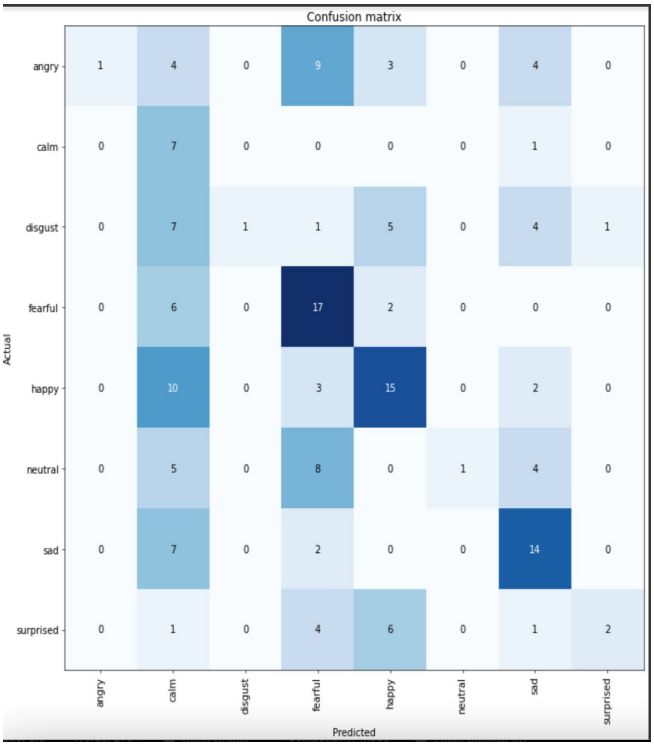
B. Results for text Recognition

After examining the input tokens using various methods, the system processes them through several layers in order to generate a response. This includes techniques like stop word elimination and other linguistic optimizations to enhance the final response. The coherence and appropriateness of the final response are closely scrutinized.To increase efficiency and response speed, the generated response is not only provided to the user but is also stored in a database. This serves two purposes. First of all, it ensures that the answer need not be written each time the same query is posed. As an alternative, the system can quickly recover the previous response from the database. This expedites the interaction procedure and lessens the quantity of redundant computational labor.



C. Results for Sentiment Analysis System

This below confusion matrix displays the value of actual and predicted values among various type of expressions.



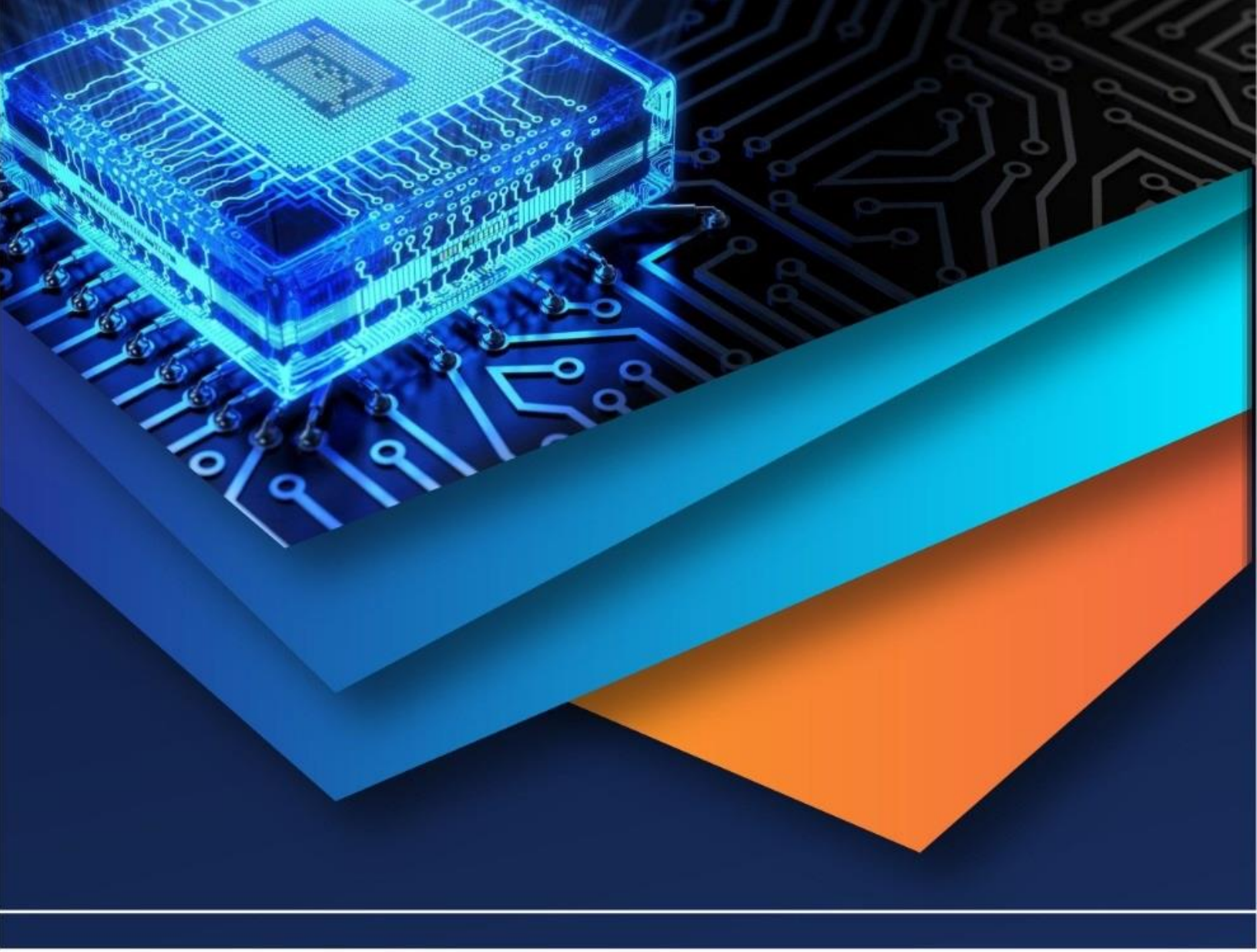
VI. CONCLUSION AND FUTURE WORK

Using AI-powered methods, models may identify user sentiments from unstructured input and categorize such sentiments as positive, negative, or neutral based on a labeled dataset. Depending on how clients' text messages are sent, chatbots can react or complete necessary duties. By applying sentiment analysis tools, you may promptly address user demands, concerns, and needs. For example, sentiment analysis can help your chatbot quickly detect a disgruntled user and route them to a live agent. Chatbots that use sentiment analysis can provide your customers with a more personalized experience.

Our methodology anticipates a seamless user experience in the future, along with steadily expanding datasets for enhanced accuracy across many platforms. Users can benefit from information retrieval in addition to practical applications like managing apps, placing calls, altering device settings, and browsing. Because of its multifunctional design, the system transforms into a versatile virtual assistant that can adapt to the changing needs of its clients. Simple voice commands enable hands-free operation, providing a user-friendly and convenient experience. This thorough integration, which not only ensures accurate responses but also streamlines tasks for optimal ease, ensures the system's relevance in users' daily lives.

REFERENCES

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
[2] Hinton, G. (2007). To recognize shapes, first learn to generate images. Progress in brain research, 165, 535-547.
[3] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[4] Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Processing Letters, 24(3), 279-283.
[5] Lim, J., Kim, T. J., & Moon, Y. H. (2018). An Audio Detection System for Misantisocial Behavior in Ambient Assisted Living. Procedia Computer Science, 125, 1023-1030.
[6] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine, 29(6), 82-97.
[7] Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 6645-6649).
[8] Schuller, B., & Batliner, A. (2014). Computational paralinguistics: emotion, affect and personality in speech and language processing. John Wiley & Sons.
[9] Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining text data (pp. 415-463). Springer, Boston, MA.
[10] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1–2), 1-135.
[11] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
[12] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.
[13] Hinton, G., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.
[14] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
[15] Xie, J., Zhu, Y., & Gao, G. (2018). Advancements and prospects of generative adversarial networks. arXiv preprint arXiv:1808.03344.
[16] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
[17] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[18] Successful audio detection in the cloud using neural networks and Purpose-built Instances. Amazon Web Services.
[19] Sak, H., Senior, A., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. Proceeding of INTERSPEECH. pp.338-342.
[20] Liu, B., & Zhang, L. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
[21] Kockmann, M., Burget, L., & Cernocky, J. (2010). Brno University of Technology System for Interspeech 2010 Paralinguistic Speech Challenge. Proceedings of INTERSPEECH. pp.2826-2829.
[22] Liu, J., & Perez, J. (2017). GANs for medical image analysis. arXiv preprint arXiv:1707.06314.
[23] Emek Yüceer, Ahmet Sonmez (2020). A Deep Learning Based Automatic Emotion Recognition System for Customer Interactions.
[24] Sentayehu, F., & Yared, H. (2019). Enhancing Deep-Learning Sentiment Analysis with Ensembles and Metaheuristics.
[25] Backouche, M., El Hassani, A.H. & Meknassi, M. (2017). Real-Time Audio Event Detection based on Classification and Regression Trees and Support Vector Machine.
[26] Nwe, T., Foo, S., De Silva, L. (2003) Speech emotion recognition using hidden Markov models. Speech Communication, 41, 603-623.
[27] Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2015). How transferable are features in deep neural networks?
[28] S. Mirjalili, S. M. Mirjalili, A. Lewis, "Grey Wolf Optimizer," Advances in Engineering Software, 2013.
[29] Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014). Facial landmark detection by deep multi-task learning. European conference on computer vision.
[30] Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning (pp. 1096-1103).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY