SYNOPSIS

On

Sentiment prediction using speech

SUBMITTED IN PARTIAL FULFILLMENT FOR THE REQUIREMENT OF THE AWARD
OF DEGREE OF

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

Submitted by
Shivaji Saxena (1900290110101)
Shailendra Pratap Singh (1900290110096)
Shalini Tyagi(1900290110097)
Abhinetra Patel(1900290210009)
Vrishank Singh(2000290110192)

Supervised by

Youddha Beer Singh

Session 2021-22

DEPARTMENT OF COMPUTER SCIENCE INFORMATION TECHNOLOGY

KIET GROUP OF INSTITUTIONS, GHAZIABAD

(Affiliated to Dr. A. P. J. Abdul Kalam Technical University, Lucknow, U.P., India)

1.Abstract

The project 'Sentiment Analysis using Voice' will help in identifying the fraud calls like scammers asking for bank details ,personal details of less educated people ,who believe these frauds easily and believe they are actually the officials from the agency ,and got scammed by them .The limitations to this can be noise and disturbance in audio data and background noises to minimize the impact of these errors we are working on a model to provide a viable solution. The project can also be used in detecting negative emotions and harmful thoughts on more research and on including some additional features .We will start from already given datasets and test the models on it after the process ,testing on data after that will start and deployment of model using a appropriate interface . The project will come in handy when used on calls for real time analysis of call , for detecting the call is fraud or valid on time.It will start from basic after learning process , understanding sound and it's features in machine from different sources , and how to handle sound data , how to preprocess it and how to build model for it.The final step involves deployment and testing of model on real voice data provided by one of our team members.

2.Introduction

- In a large proportion of these audio calls, people depict their opinions about products, movies, social issues, political issues, etc. The capability of detecting the sentiment of the speaker on calls can serve two basic functions: (i) it can enhance the retrieval of the particular data in question, thereby, increasing its utility, and (ii) the combined sentiment of a large number of audio on a similar topic can help in establishing the general sentiment. It is important to note that automatic sentiment detection using text is a mature area of research, and significant attention has been given to product reviews, we focus our attention on dual sentiment detection in calls based on audio and text analysis. We focus on audio calls because the nature of speech in these calls is more natural and spontaneous which makes automatic sentiment processing challenging.

3.Literature Review

- From paper" A Study of Support Vector Machines for Emotional Speech Recognition" In this paper, efficiency comparison of Support Vector Machines (SVM) and Binary Support Vector Machines (BSVM) techniques in utterance-based emotion recognition is studied. Acoustic features including energy, Mel-Frequency Cepstral coefficients (MFCC), Perceptual Linear Predictive (PLP), Filter Bank (FBANK), pitch, their first and second derivatives are used as frame-based features.[1]

- In paper "Audio and Text based multimodal sentiment analysis using features extracted from selective regions and deep neural networks" An improved multimodal approach to detect the sentiment of products based on their multimodality natures (audio and text) is proposed. The basic goal is to classify the input data as either positive or negative sentiment. Learning utterance-level representations for speech emotion and age/gender recognition. Accurately recognizing speaker emotion and age/gender from speech can provide better user experience for many spoken dialogue systems. In this study, we propose to use Deep Neural Networks (DNNs) to encode each utterance into a fixed-length vector by pooling the activations of the last hidden layer over time.[2]

- The paper "Towards Real-time Speech Emotion Recognition using Deep Neural Networks" proposes a real-time SER system based on end-to-end deep learning. Namely, a Deep Neural Network (DNN) that recognizes emotions from a one second frame of raw speech spectrograms is presented and investigated. This is achievable due to a deep hierarchical architecture, data augmentation, and sensible regularization. Promising results are reported on two databases which are the ENTERFACE database and the Surrey Audio-Visual Expressed Emotion (SAVEE) database.[3]

- In paper" Sentiment extraction from natural audio streams" a system for automatic sentiment detection in natural audio streams such as those found in YouTube is proposed. The proposed technique uses POS (part of speech) tagging and Maximum Entropy modelling (ME) to develop a text-based sentiment detection model. Additionally, we propose attuning technique which dramatically reduces the number of model parameters in ME while retaining classification capability. Finally, using decoded ASR (automatic speech recognition) transcripts and the ME sentiment model, the proposed system is able to estimate the sentiment in the YouTube video. In our experimental evaluation, we obtain encouraging classification accuracy given the challenging nature of the data. Our results show that it is possible to perform sentiment analysis on natural spontaneous speech data despite poor WER (word error rates).[4]

4. Feasibility Study

A feasibility study, sometimes called a feasibility analysis or feasibility report—is a way to evaluate whether or not a project plan could be successful. A feasibility study evaluates the practicality of your project plan in order to judge whether or not you're able to move forward with the project.

Feasibility studies help:

• Confirm market opportunities before committing to a project

• Narrow your business alternatives

• Create documentation about the benefits and detriments of your proposed initiative

• Provide more information before making a go/no go decision

Technical feasibility-

A technical feasibility study reviews the technical resources available for your project. This study determines if you have the right equipment, enough equipment, and the right technical knowledge to complete your project objectives.

Financial feasibility-

Financial feasibility describes whether or not your project is fiscally viable. A financial feasibility report includes a cost/benefit analysis of the project.

Market feasibility-

The market feasibility study is an evaluation of how your team expects the project's deliverables to perform in the market.

Operational feasibility-

An operational feasibility study evaluates whether or not your organization is able to complete this project.

5.Objective

- Sentiment classification is a way to analyze the subjective information in the text and then mine the opinion „Sentiment analysis is the procedure by which information is extracted from the opinions, appraisals and emotions of people in regards to entities, events and their attributes . In decision making, the opinions of others have a significant effect on customers ease, making choices with regards to online shopping, choosing events, products, entities.

- Categorization or classification of opinion sentiment into-

- · Positive

- · Negative

- Detecting Scam calls

6.Technology

- Hardware Requirements:

- · Core i5/i7 processor

- · At least 8 GB RAM

- · At least 60 GB of Usable Hard Disk Space

- Software Requirements:

- · Python 3.x

- · Anaconda Distribution

- · NLTK Toolkit

- CNN

7.Planning of Work

- DATA COLLECTION:

- Data refers to the dataset collected from a person or provided by our mentor . For training our model .

- Data preprocessing:

- Filtering data and making it suitable for training , to reduce noise and null values in dataset.

- SENTIMENT SENTENCE EXTRACTION & POS TAGGING:

- In natural language processing, part-of-speech (POS) taggers have been developed to classify words based on their parts of speech. For sentiment analysis, a POS tagger is very useful because of the following two reasons: 1) Words like nouns and pronouns usually do not contain any sentiment. It is able to filter out such words with the help of a POS tagger; 2) A POS tagger can also be used to distinguish words that can be used in different parts of speech.

- NEGETIVE PHRASE IDENTIFICATION:

- Words such as adjectives and verbs are able to convey opposite sentiment with the help of negative prefixes. For instance, consider the following sentence that was found in an electronic device's review: "The built in speaker also has its uses but so far nothing revolutionary." The word, "revolutionary" is a positive word according to the list in. However, the phrase "nothing revolutionary" gives more or less negative feelings. Therefore, it is crucial to identify such phrases. In this work, there are two types of phrases have been identified, namely negation-of-adjective (NOA) and negation-of-verb (NOV).

- Learning and Implementation of machine learning algorithms to build models to process data and predict result .

8.References

1.Sentiment Analysis of speaker specific speech data

  Author - Maghilnan S, Rajesh Kumar M, Senior IEEE, Member School of Electronic Engineering VIT University Tamil Nadu, India

https://www.researchgate.net/publication/323276680_Sentiment_Analysis_on_Speaker_Specific_Speech_Data

2. Sentiment Analysis of Speech

Author : Aishwarya Murarka , Kajal Shivarkar , Sneha, Vani Gupta ,Prof.Lata Sankpal Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India

https://ijarcce.com/upload/2017/november-17/IJARCCE%2037.pdf

3. Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network: Effects of Bandwidth Reduction and Companding-

Author : Lech M, Stolar M, Best C and Bolia R (2020) Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network: Effects of Bandwidth Reduction and Companding. *Front. Comput. Sci.* 2:14. doi: 10.3389/fcomp.2020.00014

https://www.frontiersin.org/articles/10.3389/fcomp.2020.00014/full

4. Pre-trained Deep Convolution Neural Network Model With Attention for Speech Emotion Recognition-

Author : Zhang H, Gou R, Shang J, Shen F, Wu Y and Dai G (2021) Pre-trained Deep Convolution Neural Network Model With Attention for Speech Emotion Recognition. *Front. Physiol.* 12:643202. doi: 10.3389/fphys.2021.643202

https://www.frontiersin.org/articles/10.3389/fphys.2021.643202/full

5. Stacking machine learning models for speech sentiment analysis-

Author : Zadeh A, Liang PP, Poria S, Vij P, Cambria E and Morency L-P (2018), Multi-attention recurrent network for human communication comprehension, *In Thirty-Second AAAI Conference on Artificial Intelligence.*

https://towardsdatascience.com/stacking-machine-learning-models-for-speech-sentiment-analysis-adf433488845