Research Documentation: Talk Ratio Analysis Solutions

Introduction

The objective of this research is to identify and evaluate existing solutions for Talk Ratio analysis, specifically at both an audio-based tonal level and an NLP-based level. The research assesses open-source algorithms and third-party solutions with API integrations. The aim is to determine the most suitable solution for integration into Bipper's AI-based Speech Analytics product.

Methodology

- **Identifying Solutions**: A comprehensive search was conducted to identify available solutions for Talk Ratio analysis. Emphasis was placed on solutions that cater to both audio-based and NLP-based tonal analysis.
- **Evaluation Criteria**: Solutions were assessed based on the following criteria:
- **Accuracy**: Projected accuracy in determining Talk Ratio.
- **Integration**: Ease of integration into the existing prototype.
- **Scalability**: Ability to handle large volumes of audio data.
- **Cost**: Estimated costs associated with using the solution.

Conclusion

The most suitable option for Talk Ratio analysis within Bipper's Speech Analytics product. It offers a good balance of accuracy, integration ease, scalability, and cost-effectiveness. It is essential to note that the final choice of solution should consider the specific requirements and constraints of the product, and further testing and validation should be conducted before full integration. This research documentation provides a structured overview of your findings, allowing stakeholders to understand the strengths and weaknesses of each solution and make an informed decision regarding integration into the prototype.

Prototype Documentation: Speech Analytics Prototype

Introduction

This documentation provides an overview of the Speech Analytics prototype developed for Blipper's AI-based Speech Analytics product. The prototype focuses on two key aspects: Emotion Detection and Talk Ratio Analysis. This documentation covers the approach, features, preprocessing, libraries/tools, and user manual for the prototype.

Approach

Emotion Detection:

- **Features**: Mel-frequency cepstral coefficients (MFCCs) were extracted as audio features for emotion detection. MFCCs capture audio spectral characteristics.
- **Model**: A Support Vector Machine (SVM) classifier was used for emotion classification, trained on labeled emotion data.
- **Preprocessing**: Standardization of MFCC features was performed using `StandardScaler`.

Talk Ratio Analysis:

- **Features**: The prototype uses a placeholder approach for talk ratio analysis, with a predefined talk ratio value.
- **Preprocessing**: No extensive preprocessing is performed for talk ratio analysis in this prototype.

Libraries and Tools

The prototype is built using Python and utilizes the following libraries and tools:

- **Flask**: Used for web application development and handling user interactions.
- **Librosa**: Employed for audio file loading and feature extraction.
- **scikit-learn**: Used for SVM-based emotion classification.
- **NumPy**: Utilized for numerical operations and data handling.
- **Matplotlib**: Used for generating line charts to visualize results.

Challenges Faced

- **Model Training**: Collecting and preparing labeled emotion data for training the emotion detection model was a challenging task. Finding an appropriate dataset with emotion labels was time-consuming.
- **Talk Ratio Placeholder**: Implementing a meaningful talk ratio analysis was challenging due to the lack of specific research findings. A placeholder approach was adopted for this prototype.
- **Deployment Considerations**: Preparing the application for production deployment involves additional considerations such as security, scalability, and performance optimization.

User Manual

Uploading Audio Files:

- Access the Speech Analytics web application.
- Click the "Choose File" button to select an audio file for analysis.
- Choose a valid audio file in supported formats (e.g., WAV, MP3) from your local device.
- Click the "Upload" button to submit the selected audio file for analysis.

Supported Audio Formats:

- Supported audio file formats include WAV, MP3, and other common formats compatible with the Librosa library.

Viewing Results:

- After uploading an audio file, the system will process the file for emotion detection and talk ratio analysis.
- The results will be displayed on a new page, including the predicted emotion ("happy" or "sad") and a placeholder talk ratio value.
- The predicted emotion is shown in a table with the audio file name, and a line chart displays the emotional changes over time.
- Additionally, the predicted talk ratio is presented in a table with appropriate columns, and a line chart illustrates the talk ratio trends.