REAL-TIME SPEECH-TO-SENTIMENT: SPEECH ANALYSIS USING LLMS

Aaron Park, Jeremy Ky, Davis Wang
{ync4hn,juh7hc,bqe6ue}@virginia.edu

ABSTRACT

This project aims to combine speech recognition and sentiment analysis to understand human emotions in real-time conversations. The goal of the project is to utilize state-of-the-art large language models (LLMs) for sentiment detection by analyzing transcriptions generated from speech input. Our approach leverages advanced speech recognition APIs to transcribe spoken language into text, which is then processed by sentiment analysis models such as BERT and RoBERTa, and then fine-tuned on datasets like GoEmotions. The primary objective is to assess the effectiveness of these models in accurately classifying emotions from transcribed speech, providing insights into user sentiment.

1 INTRODUCTION

As students in this NLP class, we aim to explore the intersection of speech recognition and sentiment analysis to enhance our understanding of how large language models (LLMs) perform in real-time scenarios. Specifically, we want to learn how effectively sentiment can be derived from speech transcriptions, and how state-of-the-art models like BERT and RoBERTa handle the nuances of emotional expression in spoken language. By focusing on speech-based sentiment detection, we will gain hands-on experience in fine-tuning and evaluating pre-trained models for sentiment classification tasks, a crucial skill in the field of NLP.

This project is particularly interesting because it combines two impactful areas of NLP—speech recognition and sentiment analysis—that have widespread applications, from customer service bots to mental health assistants. Real-time emotion detection can significantly enhance the interaction between users and AI, making conversational systems more empathetic and responsive.

Our timeline for the project is as follows:

- **Weeks 1-2**: Set up speech recognition APIs (Google Cloud, Whisper) and fine-tune sentiment analysis models (BERT, RoBERTa) using emotion-labeled datasets like GoEmotions.
- **Weeks 3-4**: Conduct initial testing of speech-to-text pipelines, ensuring accurate transcription for sentiment analysis. Begin evaluating the performance of sentiment analysis models on transcribed speech, focusing on basic metrics such as accuracy and precision.
- **Weeks 5-6**: Refine the sentiment detection process, improving model fine-tuning and adjusting based on feedback from initial testing. Explore more advanced sentiment metrics, including F1 score and confusion matrices, to assess model performance.
- **Weeks 7-8**: Investigate the integration of sentiment-driven response generation for potential chatbot implementation. Test how sentiment output can influence conversation flow in chatbots or assistive applications.
- **Week 9**: Finalize project, document results, and prepare for presentation. Summarize findings on the effectiveness of combining speech recognition and sentiment analysis and highlight future work possibilities, such as full chatbot integration.

By the end of this project, we expect to have a deeper understanding of how well LLMs can interpret human emotions from speech, along with practical insights into the challenges of real-time sentiment analysis.

2 Related Work

Recent studies have explored integrating voice recognition with Large Language Models (LLMs) for sentiment analysis applications. Researchers have focused on bridging the gap between audio inputs and LLMs' text-based processing capabilities. One approach, as described in "Amplifying LLMs in Emotion Recognition with Vocal Nuances" (Wu et al., 2024), involves translating speech characteristics into natural language descriptions that can be incorporated into LLM prompts. This method enables multimodal analysis without requiring architectural modifications to the LLMs. Another study, "LLM-Driven Multimodal Opinion Expression Identification" (2024), proposed a technique called STOEI that combines speech and text modalities for opinion expression identification using LLMs. These works demonstrate the potential of leveraging LLMs for more nuanced sentiment analysis by incorporating audio features, addressing the limitations of text-only approaches in capturing the full emotional context of spoken interactions.

Citations:

- https://arxiv.org/html/2407.21315v1
- https://arxiv.org/html/2406.18088v2
- https://deepgram.com/learn/ai-speech-audio-intelligence-sentiment-analysis-int
- https://www.observe.ai/blog/voice-analytics
- https://research.aimultiple.com/audio-sentiment-analysis/
- https://www.assemblyai.com/products/speech-understanding
- https://www.reddit.com/r/singularity/comments/1bpailv/you_guys_have_to_try_this_new_empathy_llm_demo_it/
- https://online.stat.psu.edu/stat414/lesson/15/15.1