Dissertation Project Description

Title: Semi-Automatic Transcription Tool with OpenAI's Whisper for Hearing Aid Research
Name: Adam Spencer
Supervisor: Prof. Jon Barker
Module Code: COM3610
Date: 03/10/2022

Introduction

The Clarity Project is a research project involving the University of Sheffield (among other universities) and aims to "stimulate progress in hearing aid signal processing by organising open machine learning challenges" (The Clarity Project, n.d.).

I have been informed by my supervisor that they have a substantial amount of data consisting of recordings of people with hearing aids repeating a phrase they have been played. This data must be accurately transcribed in order to provide benefit to the project; the previous year's data was first transcribed by a Google API and again by a group of PhD students due to a lack of confidence in the API's results. This incurred unnecessary costs to the project because the API is not free to use and nor are the students.

Recently (that is, 21/09/2022), OpenAI released an automatic speech recognition (ASR) system called Whisper to the public (Radford, et al., 2022). It has been trained on 680,000 hours of data – a huge figure for an open-source project. It is entirely free to use (unlike alternatives) and is proclaimed as state of the art.

My dissertation project is to create a software tool to semi-automate the transcription of the data collected by The Clarity Project using OpenAI's Whisper model; 'semi-automate' meaning that once Whisper has transcribed the data, a researcher will be able to correct any sections the model is not confident were correct. This will drastically reduce the human intervention required to process the data and minimise the financial cost to provide transcripts.

During this project I would also like to verify the effectiveness of Whisper for this task compared to that of the Google API.

Analysis

My first major hurdle will be using the University's high-powered computing (HPC) service to run Whisper. I am unfamiliar with cluster computing so will have to dedicate a decent chunk of time to getting this working. I have a solid understanding of Linux so I shouldn't have trouble interfacing with the computers; the concepts of cluster computing will be my focus.

I am aware that the audio recordings used by the Clarity Project were collected by the NHS and are considered sensitive. This means I will never copy the data onto a personal computer or access it outside of the Uni network.

From some reading of the Whisper GitHub page (OpenAI, 2022), I've not been able to find a simple 'confidence' metric (i.e., confidence that the transcription is correct) which would be very useful for my project. I will have to look deeper into this and perhaps create my own interface to extract this information from the model.

As for tools, I will be using a variety of Python libraries (e.g., NumPy, PyTorch, Matplotlib, Jupyter) for using Whisper & data analysis. I will also likely use some UNIX tools for data manipulation.

For developing the transcription tool, I will need to research my best options. As far as I can see, I will either use some Python GUI tools (Tkinter, Qt) or write a web app in Rust and WebAssembly. This is another major hurdle I must cross.

In order to analyse the performance of Whisper vs Google's ASR API I must familiarise myself with some data analysis methods used in ASR research. From Radford, et al. 2022, it seems word-error rate (WER) is a common metric used to compare ASR systems.

Plan of Action

My plan for semester 1 is as follows:

- For the first couple of weeks (W3, 4), get familiar with HPC and get Whisper up and running with some sample data. I don't expect this to be particularly hard, I'd just like to know what I can and can't do with HPC (e.g., doesn't seem possible to get Jupyter working on Bessemer where the data I need is stored.) I'd also like to build my understanding of Whisper itself, in preparation for doing transcription.
- For W5 and 6 I would like to do some reading on GUI development so I can start designing the tool's interface. My findings will be added to the stage 2 report.
- Week 7 will be spent doing background reading on data analysis methods.
- For W8/9 I will optimise the transcription through Whisper. This is also a good time to begin analysing the performance vs the Google API. This analysis will be written into my stage 2 report.
- For the remaining semester I will be completing the Stage 2 report.

Now, in semester 2:

- I would like to spend as much time as possible this semester working on the GUI tool. If I can complete all the work with ASR in semester 1, I will have plenty of time to focus on this.
- Writing dissertation, including GUI designs and understanding of models. Plenty of research will be required here.
- As I prefer to do design work and note taking on paper, this will also be the time to digitise these designs. This is best left for when I have the least time (i.e., nearer exam season) as it is simple relative to the rest of the project,.

Bibliography

OpenAI, 2022. *Whisper Github.* [Online]
Available at: https://github.com/openai/whisper
[Accessed October 2022].

Radford, A. et al., 2022. Robust Speech Recognition via Large-Scale Weak Supervision.

The Clarity Project, n.d. *About Us.* [Online]
Available at: https://claritychallenge.org/about
[Accessed 6 October 2022].