CHALLENGES, USE-CASE OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE IN EDUCATION

Aleksandar Jokić, Saša Adamović, University Singidunum Belgrade

I.   ABSTRACT

Speech Recognition is an exciting and fun field to get started with Machine Learning and Artificial Intelligence. This paper shows how to use ASR for educational purposes. We showed how to convert speech-to-text and Real-time recognition using Python. We wrote a program that understands what we are saying and translates it into written words. This translation is known as speech recognition. We covered work with AssemblyAI and OpenAI building a speech emotion recogniser using Python; the recognition and the training were done over static audio files, which we recorded.

Keywords: Automatic Speech Recognition ASR, Deep Learning, Speech A.I., Word error rate WER, dataset, Python, wav, audio, VSC, API, transcript

II.   INTRODUCTION

People have been talking to machines, websites and technologies for a long time. For example, "Why did you choose this path?" on their GPS devices, "Bixby, open Youtube, search for the song" on their mobile phone "Samsung". They are, at best one-way conversations unless speech recognition technology is used in one of the different ways. Misused speech recognition systems can frustrate users and alienate them. Still, it is a powerful tool that can change how people communicate with machines – thus, how users interact with products and services around them.

In every industry, at different points in the customer journey, there are opportunities to use speech technology to simplify operations, facilitate interaction with clients, and increase accessibility to a broader, diverse audience of potential new ones.

III.   WHAT IS A SPEECH RECOGNITION SYSTEM?

Speech recognition refers to technologies that identify and give meaning to human speech using large amounts of data and statistics. It functions through a system of programs and Machine Learning algorithms that interact with each other, such as pronunciation models and acoustics that "hear" and recognise spoken language, as well as language models that determine what is said in an attempt to form the most likely meaning. The technology is also known as ASR.

How does Speech Recognition work?

It works by taking human speech recordings into sets of data and breaking them down into the o most minor pieces of information, from the full address to current statements and their corresponding transcriptions. From these audio samples and text transcriptions, the technology "learns" to recognise and interpret more complex speech patterns, vocabulary and meaning. Machines also need to understand pronunciations spoken in different environments, for example, people talking in a noisy room, from the subways, buses or around a construction site or if a person is speaking from a distance and a low volume. In addition, the technology must consider variations in accent, syntax, locale expressions and ways of saying the same things within each language. Adding more subjective elements, such as intention and feel, makes it even more complicated for the speech recognition system to work in the best possible way.

History of the Speech recognition system

Computing power and Artificial Intelligence are mainly behind the advancements in this space. With massive amounts of voice data combined with faster processing, speech recognition has reached a tipping point where its capabilities are roughly equal to humans.
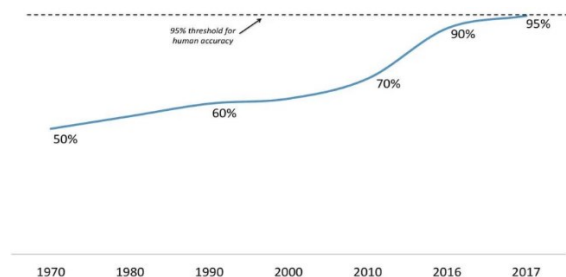
Figure 1 Mary Meeker's report on Internet trends showing word accuracy rates with A.I. speech recognition system

Although there have been many advances in recent times, voice recognition dates to the early days of the 1950s, when was announced a system of recognition numbers spoken by inventor Bell Laboratories designed the Audrey. In the 1960s and 1970s, there were many more announcements about speech recognition, where IBM introduced its own "Shoebox" that understands up to 16 words, and later "Harpy" program capable of understanding over 1000 words, which is about the same as the vocabulary of a three-year-old child. In the 1980s, we have discoveries from statistical methods, announcing Hidden Markov Model "HMM", which brings us much broader and faster possibilities to grow. Later, in the 1990s,

thanks to personal computers, rapid growth made software like "Dragon Dictate". In the last 20 years, speech came from 80% accuracy to now by last statistical data up to 95%, like the announcement of Google where their lowest error rate is 4,9%.

IV. DEEP LEARNING AND SPEECH RECOGNITION SYSTEM

The Master thesis paper reviews an overview of speech recognition technology, software development and applications. It consists of three chapters, from essential skills to advanced ones. He works with static audio files, translations and transcription and Real-time Automatic Speech Recognition. However, before we jump to the use case of this work and its implementation in real life, quality is one of the essential elements in collecting speech data. It is necessary to ensure these standards:

- Dynamic quality checks – happening during the tasks that are performed in the collection of data.
- Basic information of the truth – refers to data sets that are ground standard to measure all other datasets, resulting in what is "good" quality as a benchmark of high-quality data.
- Word error rate – the most common way to measure the performance of our ASR system
- Background noise – an ideal world for using an ASR system is a world without noise
- Industry language – to improve our Deep Neural Network and increase the number of spoken recognised, or to know the hat dataset on which we are working is at the right place.
- Speech accents and slang – variations in people's speech, different countries, cities, regions
- Data diversity is the key – more varied data, including all genders, ages, accents, and things, will improve the quality of training one ASR system.

V. THE USE CASE OF ASR IN THE MODERN WORLD OF EDUCATION

All these technologies can translate spoken language into digitalised text and analyse speech for recognition, offering us a new range of applications: devices to respond to specific voices, learning a particular users accents, phrasings, etc. Using this technology, we can help not only people with disabilities but to access learning resources for others and normalise a way to get educated and be accepted.

The most beneficial aspect of speech recognition technology is its potential to make learning at all levels accessible for those who experienced difficulty engaging with typed or written text. For example, people with dyslexia or dysgraphia frequently experience frustration with notes or typing written text. These kids are more likely to fall by the wayside and be dismissed as academically ungifted because they struggle with writing in elementary and later high school.

With a system presented in the Master's thesis, speech-to-text, on their side, these students can quickly write their thoughts and improve their possibilities during the education stages.

Machine learning and speech recognition system can be trained to understand and translate the nuances of specific voices, which bring us to the next level, that kids or students with difficulties speaking or annunciating may be close to conventional voice systems; they can talk properly again. Students and kids with motor skill disorders, long or short-term physical disabilities, and speech recognition systems increase writing production and futuristic learning.

Speech-to-text developments have been successful applications in the education sector, with their system on mobile phones, where for example, taking notes by voice can be used by educators and students simultaneously.

In the classroom uses, a benefit can be in multiple ways. For example, the speech-to-text system helps a teacher show expressions during dictation and learn more about letters and words for his students when they see writing and spelling. By watching the terms they dictate and delivering on the screen, teachers can naturally appreciate the relations between phonemes and spelling patterns. For grown students and adults, writers and learning disabilities help with surface-level pieces of writing, correct spelling and grammar, resulting from improving their capabilities of spoken language. Moreover, this system can help people with anxiety in learning, allowing them to write without worrying about spelling errors or editing.

As we mentioned before, the ASR system is growing every day. Still, there are a few unique challenges when it comes to incorporating it into advanced learning. It can take a little time to get used to using the technology. Before that time, mistranscription was more frequent, and frustrations could lead to dropping the tool before benefits were overcome. Learning to speak slowly enough to qualify written work could take time, especially when brainstorming or note-taking. Another barrier is technology hardware. For the best results, we should use the best hardware, which is, in the end, the most expensive.

VI. BENEFITS OF SPEECH RECOGNITION

Populations that may benefit from speech recognition technologies for learning include users with:

- Learning disabilities, including dyslexia and dysgraphia
- Repetitive strain injuries, such as carpal tunnel syndrome
- Poor or limited motor skills
- Vision impairments
- Physical disabilities
- Limited English Language

Benefits for students with disabilities may include improved access to the computer, increases in writing production, improvements in writing mechanics, increased independence, decreased anxiety around writing, and improvements in core reading and writing abilities.

Improved access

For students with motor skill limitations, physical disabilities, blindness/low vision, or other difficulties accessing a standard keyboard and mouse, hands-free computing through the use of speech recognition technologies may be beneficial. By removing the physical barriers to writing and navigation of the computer, you can increase student access to technology and classroom activities.

Writing production

For students with learning disabilities, speech recognition technology can encourage writing that is more thoughtful and deliberate. Studies with middle and high school students with learning disabilities have shown that input via speech is less challenging and that students frequently generate papers that are longer and better quality using speech recognition technologies.

Mechanics of writing

Speech recognition technologies, in conjunction with word processors' abilities, can help reduce some of the difficulties that students may face with writing mechanics. Because students can often write more quickly with speech recognition tools, it eliminates potential obstacles, such as difficulty with handwriting or the need to transcribe thoughts while brainstorming.

Increased independence

For students with physical disabilities, poor motor skills or learning disabilities, a human transcriber is a low-tech solution for the classroom that allows the focus to shift from the physical act of writing to expressing thoughts and knowledge. However, a transcriber makes the student dependent upon a teacher or aide for writing tasks. Students who use transcribers for writing often report "spending less time planning and organizing because they felt they were keeping the transcriber waiting, or felt embarrassed about making mistakes or asking for multiple readings of what was written."

Decreased anxiety

In addition to allowing the student to work in a more independent manner, speech recognition can allow students to write without fear of spelling errors, helping them avoid the anxieties associated with mechanics, organization, and editing; many struggling writers feel embarrassment about "the appearance of their writing due to brevity of sentence or paragraph length, illegibility of handwriting, and/or misspelt words."

VII. INSTALLATION OF REAL-TIME TRANSCRIBING APPLICATION

In order to be able to use this application, we need to go to the GitHub repo and download the application itself. To run this application, you need to have some hardware requirements like a P.C. and microphone, and from programming tools, we used ReactJS as a front-end endpoint application. AssemblyAI offers a Speech-to-text API that is built using advanced A.I. methods to transcript video and audio files. Of course, the same methods are used to build a tool for real-time speech recognition.

To install this application at your console, type:

npm install -f

After installation, it is necessary to insert your API secret key, which needs to be pasted at file server.js on line 13. Then we are ready to run and test our application. For more details regarding AssemblyAI API and documentation about that, you can find them on their website, and for testing, API is free. For more futuristic works need to have paid account to use their system. Now, when everything is installed, the application can be run with the command at the console:

npm run server
npm run client

If all pass, WebSocket will open a browser with the address: localhost:3000 and your application are ready to be used for transcription from English speech-to-text.
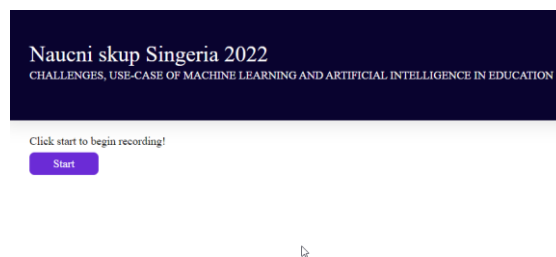
Figure 2 Testing an application for real-time transcription of speech

VIII. CONCLUSION

Reading and writing are learned skills that children acquire through time and practice. Unfortunately, this is not always possible during education time. Therefore, we should look at the Automatic Speech Recognition system as the best alternative for the tool, which can be used as a listening tool to improve the quality of education. With the ASR system, we can give a second chance and hope that people with disabilities can accomplish their wishes.

In general, this work is a report on research used as a Master thesis – Deep Learning and Speech Recognition system, and how to implement it in real life, improving day-to-day working challenges we face.

IX. REFERENCES

1. Dong Yu, Li Deng, "Automatic Speech Recognition: A Deep Learning Approach (Signals and Communication Technology)", 10. Sept. 2016,

2. Bill Gates, "Speech Recognition - The next revolution", 23 Feb. 1998,

3. K. Shalonova, "Automatic Speech Recognition", 07. Dec. 2007,

4. The complete source code of the project and subprojects can be found at the following address: https://github.com/salecivija/realtime-transcription-singeria2022.git