I'm here to propose an auto editing app

Hello, my name is Tan Zhong Ming

Just a reminder, It is supposed to be automatic

Let us do a quick recap

The first problem is remove stuttered speech manually

But this time i do a little bit modification

My second problem become improve the accuracy of subtitle

This can be done by using several techniques such as denoising, audio pre-process and deep speech

Audio pre-processing include pre-emphasis, normalize, amplify and etc. But denoising not necessary is audio pre-processing, it can be done by using AI to remove the background noise. This has been done by Nvidia, they publish a new feature for their CPU which able to remove the background noise real time by using latest RTX series GPU. Next, deep speech has a good potential become the solution. I will explain why later.

So, the scope also must change, here is my latest scope. Read slide

Why we need this app.

Now, let me share what I found on internet.

The first one of course is comparative study. Dataset always is the most important step in machine learning. No data no talk.

This is the second literature review, the paper talking about the detection and analysis of stuttered speech.

Third paper that I read is speech recognition and correction of stuttered speech.

Last one is deep speech 2. Also known as DS2. This research is done by Baidu

The goal of deep speech is to predict multiple language by only using 1 model. This is because the current ASR Automatic Speech Recognition is overcomplicated. Because current system dividing the word into phonemes, syllable and more. This will make the current system face the problem if they need to implement another new language. So, they hope to develop only one engine and applicable to all the language. And they hope to solve other issues too, such as accented speech, noisy environment and more.

This is the architecture of the DS2. Total 11 layers. They have 3 convolution layer, 7 recurrent layer and 1 fully connected layer. The fully connected layer is responsible to calculate the probability distribution and provide output to CTC layer.

The data they used for English in the experiment is WSJ, Switchboard, Fisher, LibriSpeech, and Baidu. Baidu dataset only for internal usage. So they have total eleven thousand nine hundred fourty hours of speech data. WSJ, Switchboard, Fisher you can get from Linguistic Data Consortium. LibriSpeech is the only free dataset in here. They use 95minutes data per epoch for English and 25 minute per epoch for mandarin. For data augmentation, they added noise to 40% of the dataset.

Next is the most interesting things. The read speech DS2 outperform than human. We can see some of the Word Error Rate (WER) is better than human. But accented speech is close to human performance. This can be improve by adding more data. Unfortunally noisy speech is worse than human. This is because they using synthetic voice rather than real environment recording.

The methodology that I choose is scrum. Why? Because it is fast and support incremental deliverable. Next is user story make me more understand the user requirement. Product backlog can ensure the quality of the outcome.

Software, this is important because I faced problem. This problem is not caused by human error. It is compatibility issues. The story is when I need to read the audio by using librosa library. It is a python distribution allow us access the file in the directory. “pip” and “npm” is famous and professional tools. But after I install librosa library. I cannot compile my python code. I reinstall over and over again, check through all the github stack overflow changing the version. But still didn’t solve the issues. The moment I want to reinstall the whole environment, I saw many article mentioned the issues when install the librosa by using pip. So they have manually to install the librosa, download, unzip, and put it to the folder. So I know is the distribution software problem. Then I download anaconda3, anaconda 3 also is a python distribution. Ahha, then I solve the issues, my code don’t have any problem, librosa also no problem. The problem is the python distribution. So I will use anaconda 3 as my development tools. The minimum requirement to run tensorflow is python3.5 or above. And to support tensorflow lite, you need at least support android 6.0 or API23 above.

Here is the new system workflow. It didn’t too much changes, just adding a new module. I insert deep speech module here to generate subtitle for us. Hope it can flip this industry.