Master Thesis title / MAL izenburua

Proposer(s) / Proposatzailea(k): names / izenak

Elhuyar Fundazioa

Contact / Kontaktua: email

Igor Leturia (i.leturia@elhuyar.eus)

Description / Deskribapena

We propose different master thesis projects, all ASR (Automatic Speech Recognition) related, depending on the pupil's interests:

- ASR in multilingual audios
- Speaker diarization
- Transcription of Basque dialects

Goals / Helburuak

ASR in multilingual audios

At Elhuyar we have developed two different ASR systems based on deep neural networks that work for Basque and Spanish, and each of them is able to transcribe audios, videos or live streams which contain speech in the corresponding language with very good results.

But in the Basque Country, where both Basque and Spanish are official languages, there is a great need for transcribing audios or videos that are multilingual. Precisely, sessions of the Basque Parliament, Provincial Councils and municipal plenaries are usually bilingual, with some politicians speaking in one language and others in another (depending on their political party or language knowledge). Applying any of the two transcriber systems to them leads to good results to the parts that are in that language and very bad results for the parts that are in the other language.

The work to carry out in the master thesis would be to implement different methods or approaches to transcribe multilingual audios and evaluate their results. The approaches ti try would be the following:

- Language diarization: train a neural system that will segment a multilingual audio file into different chunks according to the language spoken in it, so that then the corresponding ASR system can be applied to each chunk.
- Multilingual transcriber: train a neural system with transcriptions in both languages.
- Double transcription: apply both ASR systems to the multilingual audios and then, for each transcribed sentence, take the one for which the transcriber returns the highest probability or the one that gets the highest punctuation in a correctness score or similarity with a language model.

Speaker diarization

This project would consist of training a neural system for speaker diarization, that is, for segmenting an audio or video into chunks depending on speaker changes, and recognizing the chunks with the same speaker. That way, the transcription can also contain metadata describing who is saying each chunk.

Transcription of Basque dialects

The ASR system for Basque that we have developed at Elhuyar is able to transcribe speech in standard unified Basque, but does not work so well with the various dialects of Basque. This project would consist of training a neural network for transcribing a Basque dialect (Gipuzkoan, Biscayan, Navarro-Lapurdian, Souletin). In order to do that, a corpus of texts of that language would have to be compiled for the language model and, if necessary, also audios and transcriptions of that dialect in order to train the acoustic model.

Requirements / Betebeharrak

Computer programming skills will be needed, preferably Python. Experience in training and using deep neural network systems would be interesting. The Kaldi ASR toolkit will be used. The working environment will be Linux.

Tasks and plan / Atazak eta plana

ASR in multilingual audios

- T1.1: prepare multilingual training audios combining already available monolingual audios.
- T1.2: train a language diarization system.
- T2: train a multilingual transcriber using the transcriptions and text corpora that have been used for training both monolingual transcribers.
- T3: develop and try different systems for evaluating the correctness of the transcribed sentences.
- T4: evaluation of the three systems.

Speaker diarization

- T1: prepare training set, combining the available training sentences and their transcriptions used for developing the ASR systems.
- T2: training a neural network system.
- T3: evaluation of the results.

Transcription of Basque dialects

- T1: collecting text corpora for the language model.
- T2: evaluate results of transcriber with new language model.
- T3: if necessary, prepare a corpus for training the acoustic model also.
- T4: evaluate results of whole system.