Deep learning ASR-based for non-native learner mispronunciation detection

Toan Pham Van
Framgia Inc R&D Group
pham.van.toan@framgia.com

Thanh Ta Minh
Le Quy Don Technical University
ta.minh.thanh@framgia.com

Hau Nguyen Thanh
Framgia Inc R&D Group
nguyen.thanh.hau@framgia.com

***Abstract*—Mispronunciation and accent detection based on computer techniques are receiving increased attention nowadays. It is one of the most important part of second language acquisition. It can help non-native speakers to identify errors, learn sounds and vocabulary, and improve pronunciation performance. Deep learning is a technique used a lot recently. There are several methods of deep learning that we use there. Recurrent Neural Network (RNN) is a class of artificial neural network where connections between units form a directed graph along a sequence. This allows it to exhibit dynamic temporal behavior for a time sequence. We use RNN with 3 layers of LSTM cell. Another deep neural network class we use there is Convolutional Neural Network (CNN), it also find the connectivity between features and help us to improve the accuracy. The architecture of CNN we use there is combination of convolutional layers, maxpool and fully connected layers. Finally we use both RNN and CNN in a combination called deepspeech, they help us increase the accuracy in this works.**

Keywords*—Pronunciation Detection, Phonetic Classification, Speech Recognition, Deep Learning Speech

I. Introduction

Miliions of people over the world studying at least a foreign language. However, many of them can't approach to method for master proper pronunciations. Nowadays, with the powerful of computer system and internet, an automated computer tools to help language learners is really needed. It try to detect mistakes made by non-native learners at either the phone, word, or sentence and inform the learner of those errors. Our approach is to detect phone-level mispronunciations in words detected by an Automatic Speech Recognition (**ASR**) system. The combined approach neight CNN and RNN was constructed and and we chosen a language not popular in previous researches is Japanese as an experiment for our solution.

II. Related works

A. Feature Extraction

As discussed above, phone-level mispronunciation detection can detect mispronunciations in units of phones, words or sentences. Firstly to do it, the speaker's speech samples are first converted to certain types of features such as Linear predictive cepstral coefficients (**LPCC**), Mel-frequency cepstral coefficients (**MFCC**), Power spectral analysis (**FFT**) or Mel scale cepstral analysis (**MEL**) etc. This features used as the input of classifier. This allows the system to classify the mispronunciation by types. In our works, we using **MFCC** method with 13 features chosen.

B. Dataset

C. DNN-HMM methods

D. Result

III. Conclusion and future works

Application of Research

The research result was applied in **Chatty Pheasant Applicaiton** - a service to help non-native learner improve their Japanese pronunciation of ***Framgia Inc***[1]

Acknowledgment

This research was partially supported by ***Framgia Vietnam***. We are thankful to our colleagues who provided expertise that greatly assisted the research, although they may not agree with all of the interpretations provided in this paper.

References

[1] BLOM, ALEXANDER, and SOFIE THORSEN. "A sentiment-based chat bot." (2013).

[1] www.recruit.framgia.vn