Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Ye Kyaw Thu
Assignment title:	NLP Paper
Submission title:	asr4burmese
File name:	asr2021.pdf
File size:	245.23K
Page count:	6
Word count:	4,034
Character count:	20,845
Submission date:	27-Jan-2021 10:52PM (UTC+0630)
Submission ID:	1495359481

Automatic Speech Recognition System for Burmese Sentences using Kaldi

Hlaing Myat Nwe
hlaingmyatnwe@utycc.edu.mm

Khant Khant Win Tint
khantkhantwintint@utycc.edu.mm

Khaing Hsu Wai
khainghsuwai@utycc.edu.mm

***Abstract*—The accuracy of automatic speech recognition (ASR) remains one of the most important research challenges based on the vocabulary size, noise and the variety of language and speaker. This project is aimed to built an accurate automatic speech recognition system for small amount of data-set using Kaldi, an open-source toolkit for speech recognition written in C++. In this project, we made three types of experiment. First of all, we applied the incremental training concept and different approaches of training and adaptation techniques are studied in order to improve the recognition accuracy. Second, we used different n-gram language models to investigate the accuracy and evaluated the performance of the model depending on the different model size. For training the acoustic part of the model, Hidden Markov Model and Gaussian Mixture Model is used. The performance of the system is evaluated in terms of Word Error Rate (WER).**

***Index Terms*—Automatic Speech Recognition, Acoustic Model, Kaldi, Hidden Markov Model, Gaussian Mixture Model**

I. Introduction

Speech is the general communication language for humans. In speech processing, automatic speech recognition (ASR) is a process that converts human speech to a sequence of words which is spoken by human. The use of speech for interacting with the computer may assist the developing country as the language technologies being implemented for the e-governance system. There is only a few of works have been done in ASR for Myanmar language. The major difficulty in the research process of Myanmar language ASR is the lack of Myanmar speech corpus. Generally, it is not easy to build the speech corpus because it requires a huge amount of speech data, time, and efforts.

The main problem of the ASR is the complexity of the human language. In speech recognition system, many parameters may also affect the performance of recognition such as the vocabulary size, noise and the variety of language and speaker. In this project, different approaches have been tried to get good performance of the speech using different parameters.

The main purpose of this project is to develop more efficient acoustic model that is one of the main components of the ASR to get the better accuracy of small Myanmar ASR model. Better accuracy can be achieved if the efficient acoustic modeling approach is applied. In this project, Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) based acoustic model is built for developing the Myanmar ASR using Mel Frequency Cepstral Coefficient (MFCC) feature extraction technique. The hyperparameters of GMM-HMM are optimized to improve the ASR performance for Myanmar language. In this work, we used SRILM to build language model for Myanmar language and weighted finite-state transducers for decoding ASR. In our experiment, we made three types of evaluation based on the amount of training data, language model and number of model size in terms of WER%. We also applied and compared different approaches of training and adaptation techniques. Furthermore, We have applied myG2P mapping to create a lexicon file [1].

Among many toolkits for the implementation of speech recognition, we used Kaldi toolkit, an open-source toolkit made for dealing with speech data in our project. It is written in C++ and licensed under the Apache License v2.0 [2].

II. Data Preparation

Data preparation is a necessary step to set up ASR system with our own data. There are three parts to prepare: audio data, acoustic data and language data.

A. Audio Data

Firstly, we prepared audio data to set up an ASR system based on that data. We recorded the audio files with the students of NLP class in the E-learning studio room of University of Technology (Yatanarpon Cyber City). File format that we used is WAV. The set parameters of audio are mono channel, 16kHz sampling frequency rate and the record second is 3. Each file contains one short sentence. Each of these audio files is named in a recognizable way and placed in the recognizable folder representing particular speaker. In this project, we prepared four types of dataset for the incremental training such as we used 6 speakers for first experiment, 10 speakers for second experiment, 16 speakers for third experiment and 20 speakers for fourth experiment. The audio dataset used for our experiment is described as shown in Table I.

B. Acoustic Data

We created some text files to communicate with our audio data. In this project, we collected 500 general sentences for each speaker. Example sentences can be seen in