Speech to Intent Mapping System For Low Resourced Languages

Yohan Karunanayake

188084V

Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science (Research) in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

January 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters thesis/dissertation under my supervision.

Name of the Supervisor: Dr. Uthayasanker Thayasivam
Signature of the Supervisor: Date:

Name of the Supervisor: Dr. Surangika Ranathunga
Signature of the Supervisor: Date:

ACKNOWLEDGEMENTS

It would never be possible to finish my dissertation without the encouragement, support, and supervision of various personalities, including my mentors, my friends, colleagues, and my family. At the end of this thesis, I would like to thank all those people who made this achievable and memorable experience for me.

First and foremost, I would like to thank my supervisors Dr. Uthayasanker Thayasivam and Dr. Surangika Ranthunga for the continuous support and guidance I received in every aspect while completing this research.

I would also like to thank my progress review committee, Dr. Peshala G. Jayasekara, and Dr. Charith Chitraranjan for their valuable insights and guidance. Their advice helped me to improve the state of my research work.

This work was funded by a Senate Research Committee (SRC) Grant of the University of Moratuwa. Further, I would like to thank LK Domain Registry for the support given for conference participation via travel grants.

Finally, I would like to express my sincere gratitude all of my friends. Not just academic work, I was able to enjoy my time while involving different activities and outings. It helped me to keep my life balanced. I thank my parents who have given me a very fortunate life and always believing, trusting and supporting me.

ABSTRACT

Today we can find many use cases for content-based speech classification. These include speech topic identification and speech command recognition. Among these, speech command-based user interfaces are becoming popular since they allow humans to interact with digital devices using natural language. Such interfaces are capable of identifying the intent of the given query.

Automatic Speech Recognition (ASR) sits underneath all of these applications to convert speech into textual format. However, creating an ASR system for a language is a resource-consuming task. Even though there are more than 6000 languages in the world, all of these speech-related applications are limited to the most well-known languages such as English, because of the high data requirement of ASR. There is some past research that looked into classifying speech while addressing the data scarcity. However, all of these methods have their limitations.

This study presents a direct speech intent identification method for low-resource languages with the use of a transfer learning mechanism. It makes use of three different audio-based feature generation techniques that can represent semantic information presented in the speech. They are unsupervised acoustic unit features, character and phoneme features. The proposed method is evaluated using Sinhala and Tamil language datasets in the banking domain. Among these, phoneme based features that can be extracted from Automatic Speech Recognizers (ASRs) yield the best results in intent identification. The experiment results show that this method can have more than 80% accuracy for a 0.5-hour limited speech dataset in both languages.

Keywords: Speech Intent Identification, Spoken Language Understanding, Low-Resource Languages.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AM	Acoustic Model
AMDTK	Acoustic Model Discovery Toolkit
ANN	Artificial Neural Network
ASR	Automatic Speech Recognition/Recognizer
CNN	Convolutional Neural Networks
CTC	Connectionist Temporal Classification
DBN	Dynamic Bayesian Network
DNN	Deep Neural Network
FNN	Feed-forward Neural Networks
GMM	Gaussian Mixture Models
GPU	Graphics Processing Unit
HLT	Human Language Technologies
HMM	Hidden Markov Model
LM	Language Model
LSTM	Long Short Term Memory
LVCSR	Large Vocabulary Continuous Speech Recognition/Recognizer
MFCC	Mel Frequency Cepstral Coefficients
NLU	Natural Language Understanding
RNN	Recurrent Neural Network
SVM	Support Vector Machine
WER	Word Error Rate

TABLE OF CONTENTS