

**DOMAIN SPECIFIC VOICE INTENT CLASSIFICATION WITH
BLSTM**

Hellarawa Mudiyansele Madushika Jayani Hellarawa

(199325F)

Degree of Master of Science in Computer Science

Department of Computer Science and Engineer

University of Moratuwa

Sri Lanka

October 2022

DOMAIN SPECIFIC VOICE INTENT CLASSIFICATION WITH BLSTM

Hellarawa Mudiyanseelage Madushika Jayani Hellarawa

(199325F)

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

October 2022

DECLARATION

I declare that this is my own work and this thesis/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.

I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: *UOM Verified Signature*

Date: 16.10.2022

Name: H.M.M. Jayani Hellarawa

The above candidate has carried out research for the Master's thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Signature: *UOM Verified Signature*

Date: 16.10.2022

Name:

Dr.T.Uthayasanker

ACKNOWLEDGEMENTS

I would like to express my appreciation and the gratitude for those of who supported me throughout process as this research would not be possible without those. My sincere gratitude goes to my supervisor Dr. Uthayasanker Thayasivam for his continuous support and guidance throughout. And also, for all the academic and non-academic staff, Department of Computer Science and Engineering, Faculty of Engineering, University of Moratuwa Sri Lanka for their support.

My sincere thanks should go to all my friends and colleagues who helped me to make this a success.

Finally, I must express my profound gratitude to my parents, friends and colleagues who helped me to make this a success by providing me with unfailing support and continuous encouragement. This accomplishment would not have been possible without them. Thank you.

Author
Jayani Hellarawa

ABSTRACT

With the current global pandemic all countries around the globe are facing difficulties managing their healthcare services in a way that ensures the high availability of critical services while maintaining the safety of both the patient and the staff. According to Gartner's top 10 strategic technology trends 2021 [1], it says "*Rather than building a technology stack and then exploring the potential applications, organizations must consider the business and human context first.*" where it highlights the need for human centric development while stating that it is the IT leaders that decides what combination of the trends to involve in driving the most innovation and strategy.

A decade ago, simply having a website was enough to impress prospective customers and help them find their way to a service or information need and to establish a brand loyalty or identity. The growth of the technology is demanding more innovative strategies to adopted to every small to large industries that are at any stage of maturity of their roadmap to success. The increasing demands of the clients and the ability to keep a loyal customer base has highlighted the need of having a more natural way of handling a customer's inquiry gives a competitive advantage for any business.

The disappointment due to a customer getting added to a call waiting queues to reach a particular service is very critical and can even cause a loss of business opportunity. Understanding call intents can help a service provider to adapt the business engagement with the outside in a way that customers are positively satisfied which could in return increases the sales revenue. Not only that, but indirectly enables the ability for business to allocation agents or help-desk staff optimally thus avoid understaffing and overstaffing situation, which are indirect costs for any revenue-based figure.

Automation is where the technology is used to automate tasks that once required humans. Here, the menu-based call center automations can be taken as a replacement to the legacy call center agent where the human tasks were replaced by automation. The concept of hyperautomation is where the businesses are rapidly adopting it's revenue-based processes and IT process for automation. And the current state-of-the-art deals with lot of advanced technologies like Machine Learning (ML) and Artificial Intelligence (AI). Where AI and ML are used for extending the capabilities of automations.

The building of a speech recognition (ASR) systems for an open domain has been research for a lone time. Where the most of those are accomplished by collecting the voice corpus, convert them into text and performing a text classification on top of the converted text. However, this comes with lot of limitations, thus is not identified as the most feasible way of deriving intents of a speech query for a specific domain [2]. Therefore, in this research, that is focused on domain specific voice intent classification will be aligned with the healthcare domain for the English language based on a neural network with Bidirectional Long Short-term Memory (BLSTM).

TABLE OF CONTENTS

| | |
|---|-----|
| Declaration | I |
| Acknowledgements | II |
| Abstract | III |
| Table of contents | V |
| List of figures | VII |
| List of tables | IX |
| List of abbreviations | X |
| 1. Introduction | 1 |
| 1.1. Research problem | 3 |
| 1.2. Motivation | 4 |
| 2. Literature review | 5 |
| 2.1. Speech recognition | 5 |
| 2.2. Speech recognition techniques | 6 |
| 2.2.1. Hidden markov model | 8 |
| 2.2.2. Neural networks (nn) | 8 |
| 2.3. Bidirectional long short-term memory (blstm) | 8 |
| 2.4. Speech feature extraction | 9 |
| 2.4.1. Mel frequency cepstral coefficients (mfcc) | 10 |
| 2.4.2. Connectionist temporal classification | 11 |
| 2.5. End-to-end learning | 12 |
| 2.6. Transfer learning | 12 |
| 2.7. Data augmentation | 15 |
| 2.8. Healthcare domain | 17 |
| 2.9. Data collection methods | 20 |
| 3. Proposed solution | 22 |
| 4. Implementation | 24 |
| 4.1. Domain identification | 24 |
| 4.1.1. Identification of intents | 24 |
| 4.1.2. Identification of inflections | 30 |

| | |
|---|-----------|
| 4.2. Data collection | 33 |
| 4.3. Data preprocessing | 36 |
| 4.4. Benchmark and proposed architecture | 40 |
| 4.5. Model training | 41 |
| 5. Discussion | 45 |
| 6. Conclusion | 48 |
| 7. References | 50 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1: Acoustic model training process | 6 |
| Figure 2: Architecture of a asr with HMM and GMM | 7 |
| Figure 3: Structure of RNN and BNN | 9 |
| Figure 4: LSTM cells and it's operations | 9 |
| Figure 5: BLSTM architecture | 9 |
| Figure 6: MFCC windowing | 10 |
| Figure 7: Block diagram of MFCC | 11 |
| Figure 8 : Learning process of traditional ML | 13 |
| Figure 9 : Learning process of transfer learning | 13 |
| Figure 10: An overview of different settings of transfer | 14 |
| Figure 11: For fine tuning | 14 |
| Figure 12: As a feature extractor | 14 |
| Figure 13: Augmentation with time shifting | 15 |
| Figure 14: Augmentation with pitch changing | 16 |
| Figure 15: Augmentation by changing the speed | 16 |
| Figure 16: Augmentation by noise injection | 17 |
| Figure 17: Query and the concept transitions. | 18 |
| Figure 18: Concept transition inference | 19 |
| Figure 19: User basic information and overall experience | 25 |
| Figure 20: Gender distribution - participants | 25 |
| Figure 21: Age distribution - participants | 26 |
| Figure 22: Healthcare access method distribution of the survey | 26 |
| Figure 23: User experience on menu-base systems | 27 |
| Figure 24: User preference for an operator assistant over the menu-based selection | 27 |
| Figure 25: Most frequent queries received at a call-center. | 28 |
| Figure 26: Most important queries received at a call-center. | 29 |
| Figure 27: User inputs for individual intents | 30 |
| Figure 28: Data collecting web application architecture | 33 |
| Figure 29: Web application for data collection | 34 |
| Figure 30: Record inflections | 35 |
| Figure 31: Gender distribution of the audio clips | 36 |

| | |
|--|----|
| Figure 32: Audio clips by gender and validity | 36 |
| Figure 33: Invalid clips by the reason | 37 |
| Figure 34: Audio clips distribution by the intent | 37 |
| Figure 35: Number of clips by duration | 38 |
| Figure 36: Proposed model architecture | 40 |
| Figure 37: Benchmark architecture | 40 |
| Figure 38: Accuracy over batch size | 41 |
| Figure 39: Training history of different epochs; 50, 150 and 250 | 42 |
| Figure 40: Final model architecture | 43 |

LIST OF TABLES

| | |
|--|-----|
| Table 1: Most common intents | 31 |
| Table 2: List of inflections for each intent | 311 |
| Table 3: Dataset duration | 38 |
| Table 4: Accuracy over batch size | 41 |
| Table 5: Model results summary | 42 |
| Table 6: Results summary in healthcare dataset | 43 |
| Table 7: Comparison of dataset of banking domain and healthcare domain | 44 |
| Table 8: Performance comparison on audiomnist dataset | 44 |

LIST OF ABBREVIATIONS

| | |
|-------|---|
| ASR | - Automatic Speech Recognition |
| BLSTM | - Bidirectional Long Short-term Memory |
| LSTM | - Long Short-term Memory |
| ROI | - Return on Investment |
| SVM | - Support Vector Machine |
| CTC | - Connectionist Temporal Classification |
| DNN | - Deep Neural Network |
| DCT | - Discrete Cosine Transform |
| MFCC | - Mel-Frequency Cepstral Coefficient |
| RASTA | - Relative Spectral Amplitude Filtering |
| PLP | - Perceptual Linear Predictive |
| LPC | - Linear Predictive Coding |
| WFST | - Weighted Finite State Transducer |
| AM | - Acoustic Model |
| LM | - Language Model |
| SVM | - Support Vector Machine |
| FFT | - Fast-Fourier-Transformation |
| TL | - Transfer Learning |
| ML | - Machine Learning |
| AI | - Artificial Intelligence |
| LRL | - Low Resource Languages |
| HIPAA | - Health Insurance Portability and Accountability Act |