Continuous Bengali Speech Recognition Based On Deep Neural Network

Md Taohidul Islam, Arnab Sen, Sajib Kumar Chowdhury and Md Sams Noor Tonmoy

September 27, 2023

1 Introduction

The field of Automatic Speech Recognition (ASR) has witnessed significant advancements, particularly with the advent of Deep Neural Networks (DNNs). This thesis proposal aims to investigate and benchmark the performance of DNN-HMM (Hidden Markov Models) and GMM-HMM-based acoustic models for the task of continuous Bengali speech recognition. The primary objective is to assess the effectiveness of DNN-based models in comparison to traditional GMM-based models using the SHRUTI corpus, which comprises 21.64 hours of Bengali speech data.

2 Research Objectives

1. To implement DNN-HMM and GMM-HMM-based acoustic models for Bengali speech recognition.

2. To train and evaluate these models using the SHRUTI corpus.

3. To benchmark the performance of DNN-HMM and GMM-HMM-based models in terms of recognition accuracy, speed, and robustness.

4. To identify the strengths and weaknesses of each approach and conclude the suitability of DNN-based models for Bengali speech recognition.

3 Literature Review

- Review literature on automatic speech recognition, with a focus on DNN-based approaches and their applications in various languages.

- Explore previous Bengali speech recognition studies, including traditional and modern methods.

- Discuss the challenges specific to Bengali speech recognition and the potential advantages of DNNs in addressing these challenges.

4 Research Questions

1. How do DNN-HMM-based acoustic models compare to GMM-HMM-based models in terms of recognition accuracy for continuous Bengali speech?

2. What is the impact of model complexity on recognition performance?

3. How do DNN-based models perform in handling variations in Bengali pronunciation and dialects?

4. Are there specific contexts or conditions where one model type outperforms the other?

5 Methodology

1. **Data Preprocessing:** Prepare and preprocess the SHRUTI corpus for training and testing. This includes data cleaning, feature extraction, and alignment.

2. **Model Implementation:** Implement both DNN-HMM and GMM-HMM-based acoustic models for Bengali speech recognition.

3. **Training and Evaluation:** Train the models using appropriate training algorithms and evaluate their performance using a variety of metrics, including Word Error Rate (WER) and accuracy.

4. **Comparison:** Conduct a detailed comparative analysis of the DNN-HMM and GMM-HMM models, considering factors like recognition accuracy, speed, and robustness.

5. **Discussion:** Discuss the findings and draw conclusions regarding the effectiveness of DNN-based models for Bengali speech recognition.

6 Significance and Contribution

This research is significant as it contributes to the understanding of the suitability of DNN-based models for Bengali speech recognition. The findings will provide insights into the potential advantages and limitations of these models and may inform future developments in ASR technology for Bengali and other languages.

7 Timeline

- Data Collection and Preparation: 6 months

- Model Development and Training: 12 months

- Evaluation and Analysis: 3 months

- Thesis Writing and Finalization: 6 months