Transformer Based End-To-End Speech Recognition System For Bangla Language

Submitted By :
MD RIFAT HOSEN

Supervised By :
Dr. SANGEETA BISWAS

End-To-End Speech Recognition

আমি ভালো আছি.....

Speech

Text

Problem Statement

Figure: Regional Bangla Speech-to-Text Error.

Previous Work on Bangla

Methods:

1. **SVM** (Support Vector Machine)
2. **CNN** (Convolutional Neural Network)
3. **CTC** (Connectionist Temporal Classification)
4. **HMM** (Hidden Markov Model)

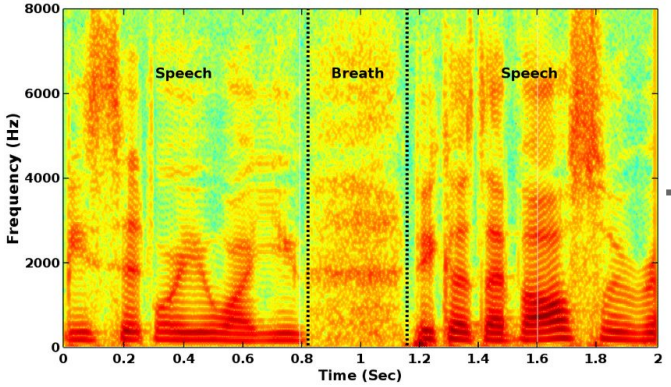
Models:

1. **BanglaASR**

New Developments In Our Work

1. **Using Transformer Based Model**
2. **Using Large Data Set**
3. **Using Regional Dialect**

Speech Transformer



Speech Feature

convolutional Layer

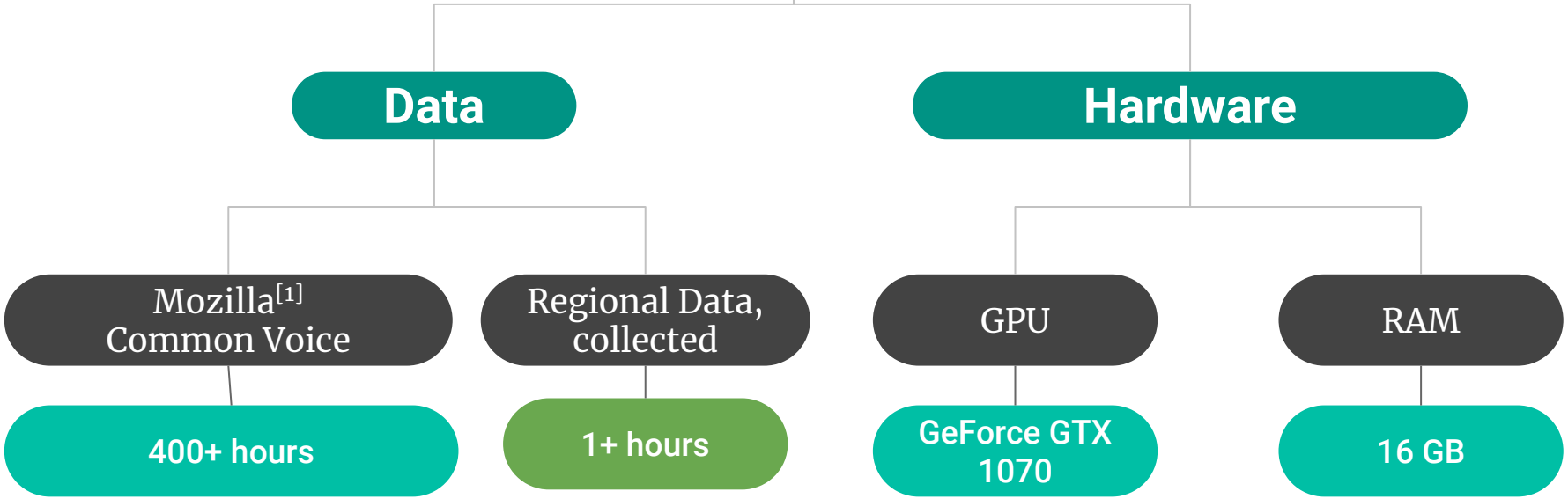
Encoder Block

Character Features

Decoder Block

Output
(Text)

Resource


Data
Hardware
Mozilla[1] Common Voice
Regional Data, collected
GPU
RAM
400+ hours
1+ hours
GeForce GTX 1070
16 GB

Base Model Output

1. CTC Model

Data: 20 K / 35 Hr

Training Time: 8 Hr

```
--------------------------------
Target    : এইসব লিপি তখনকার মানুষের উন্নত শিল্পজ্ঞান নির্দেশ করে।
Prediction: এই সবিপি তখনকার মানাষের নন্যত শিলক ডো নের দেশ করে।
--------------------------------------------------------------------------------
--------------------------------
Target    : এই কাজের জন্য বিশেষ পারদর্শিতা থাকা শিল্পী থাকে।
Prediction: এই কাজের জন্য বিশেষ পারদরশিডা থাকা সিল্পি ঢাকে।
--------------------------------------------------------------------------------
```

Accuracy - 58%

Trained Model Output

2. Transformer Model

Data: 30 K / 55 Hr

Training Time: 15 Hr

```
--------------------------------
Target     : তিনি বিবাহিত এবং তাঁর চারটি সন্তান রয়েছে।
Prediction: তিনি বিবাহিত এবং তার চাটি সহন্তান রয়েছে।
--------------------------------------------------------------------------------
--------------------------------
Target     : কিন্তু এখনো দারিদ্র্য বিদ্যমান।
Prediction: কিন্তুও এখন দারিক গ্রবিত মা কেন।
--------------------------------------------------------------------------------
--------------------------------
Target     : তিনি ক্যাম্পের কমান্ডারের বক্তব্য তুলে ধরেন।
Prediction: তিনি কেনদের কমান্ডারের বক্ক ব দুলেদ করেন।
```

Accuracy - 68%

Previous Works on Bangla

Author	Data Type	Approach	Accuracy
Noman et al. in 2022	Words	ANN	95.23%
Sen et al. in 2021	Digits	CNN	96.7%
Swarna et al. in 2020	sentence	SVM	53.75%

User Interface for ASR Model

CSE-RU Bangla Speech Recognizer

Translate Bangla Speech to Text.

Start Voice Input

Choose File | No file chosen | Upload

Speech To Text | Convert To Text

Some Sample of Collected Data

চাঁপাইনবাবগঞ্জ

মনে মনে কি বলছ?

মনে মনে কি কহিছো?

আমি একটু একটু বুঝি।

হামি এল্ল্যা একটু হইলেও বুঝি।

আমার বমি বমি লাগছে।

হার বমি বমি লাগছে।

আমি রাগ করেছি।

হামি রাগ কইর‍্যাছি।

আমার ভালো লাগছে না।

হার ভালো লাগছে না।

চট্টগ্রাম

আমার স্থায়ী ঠিকানা চট্টগ্রাম

ম স্থায়ী ঘর চট্টগ্রাম

আমি রাজশাহী বিশ্ববিদ্যালয়ের বাংলা বিভাগে পড়াশোনা করছি

মুই রাজশাহী বিশ্ববিদ্যালয়ত বাংলা বিভাগেত পড়াশোনা গরঙর

আমার সঙ্গে কথা বলিও না তো

ম লগে হদা নহোজ দে

আমরা হয়ত কিনছি

আমি হিনিত্তেই আয়

ওদের জন্য এটা কতটা কষ্টের ব্যাপার

তারাত্তেই ইয়ান হত্তমান হষ্টর ব্যাপার

Some Sample of Collected Data

চাকমা

আমি বাজি ধরে বলতে পারি এই পানিতে চিংড়ি আছে।

মুই বাজি গরি হোই পারং এ পানিয়ানদ ইজে আগন।

হাত নিচে রাখ!

আত্তানি তলে রাগা!

আমি সবসময় তাকে বলতাম আমি ঠিক আছি।

মু ই তারে যেক্কে অদসাদ হঙ্গে মুই গম আগং।

আমি কথা দিয়ে কথা রাখি।

মুই বাজ দিলে বাজ রাগাং।

Our Collected Dataset

Dialect	**Unit**
Chakma	5871 Sentences
Chapai	2870 Sentences
Chatgaiya	4913 Sentences
	Total : 13654 Sentences

Future Work

Bangla speech to Text Translation (**local dialect**).

Building a dataset for improving system.

Publish Our works on transactions on asian and low-resource language information processing Journal

Constraint

- Computational Power (**GPU,TPU**)
- Time Consuming Pre Process Stage(**Data Annotation**)
- Data collections(**Data Privacy**)

Reference

1. https://commonvoice.mozilla.org/en/datasets#other-datasets [Dataset]
2. https://ieeexplore.ieee.org/abstract/document/8457100 [Paper]
3. http://dspace.bracu.ac.bd/xmlui/handle/10361/13632 [Paper]
4. https://www.academia.edu/22666632/Instant_Bangla_Speech_to_Text_Conversion [Paper]
5. https://www.researchgate.net/publication/277671513_Implementation_of_speech_recognition_system_for_Bangla [Paper]
6. https://keras.io/examples/audio/transformer_asr/ [Resource]

Thank You