The Makerere Radio Speech Corpus: A Luganda Radio Corpus for Automatic Speech Recognition

MAKERERE UNIVERSITY

Makerere AI Lab

Jonathan Mukiibi, Andrew Katumba, Joyce Nakatumba-Nabende, Ali Hussein, Josh Meyer

jonathan.mukiibi@students.mak.ac.ug,{andrew.katumba, joyce.nabende}@mak.ac.ug, alizawahry1@gmail.com, josh@coqui.ai

Introduction

In sub-Saharan Africa, low internet penetration makes radio the most preferred medium of social communication. Radio provides an opportunity for people's concerns, particularly in rural communities, to get heard through the various radio talk shows where they can call in. Radio creates a unique platform where views that could potentially harness the development of better policies are discussed.

In order to achieve this, there is need to create high quality speech datasets that can be used to build usable radio monitoring automatic speech recognition (ASR) systems for under-resourced languages.

In this paper, we describe the process of creating a radio speech corpus for Luganda, a Low resourced language. To our knowledge, this is the first publicly available radio dataset in sub-Saharan Africa.

We present baseline Luganda ASR performance results using Coqui STT toolkit, an open source speech recognition toolkit. We release 155 hours of The Makerere Radio Speech Corpus.

Main Contributions

- Methodology used to collect and create the Luganda radio speech corpus.
- Openly release 155 hours of the radio dataset.
- Present the first radio monitoring Connectionist Temporal Classification (CTC) end-to-end ASR model for Luganda
- Evaluate the performance of the ASR model on a COVID-19 radio conversation test set
- We show how hotword boosting can improve keyword detection in a COVID-19 use case-based radio monitoring system
- Evaluate the performance of the Luganda Radio ASR on gender.

Related Work

- Previous work in the area of radio browsing using Automatic Speech Recognition (ASR) has been done by the United Nations (Menon et al., 2018a)
- Common Voice Dataset

Radio Speech Corpus

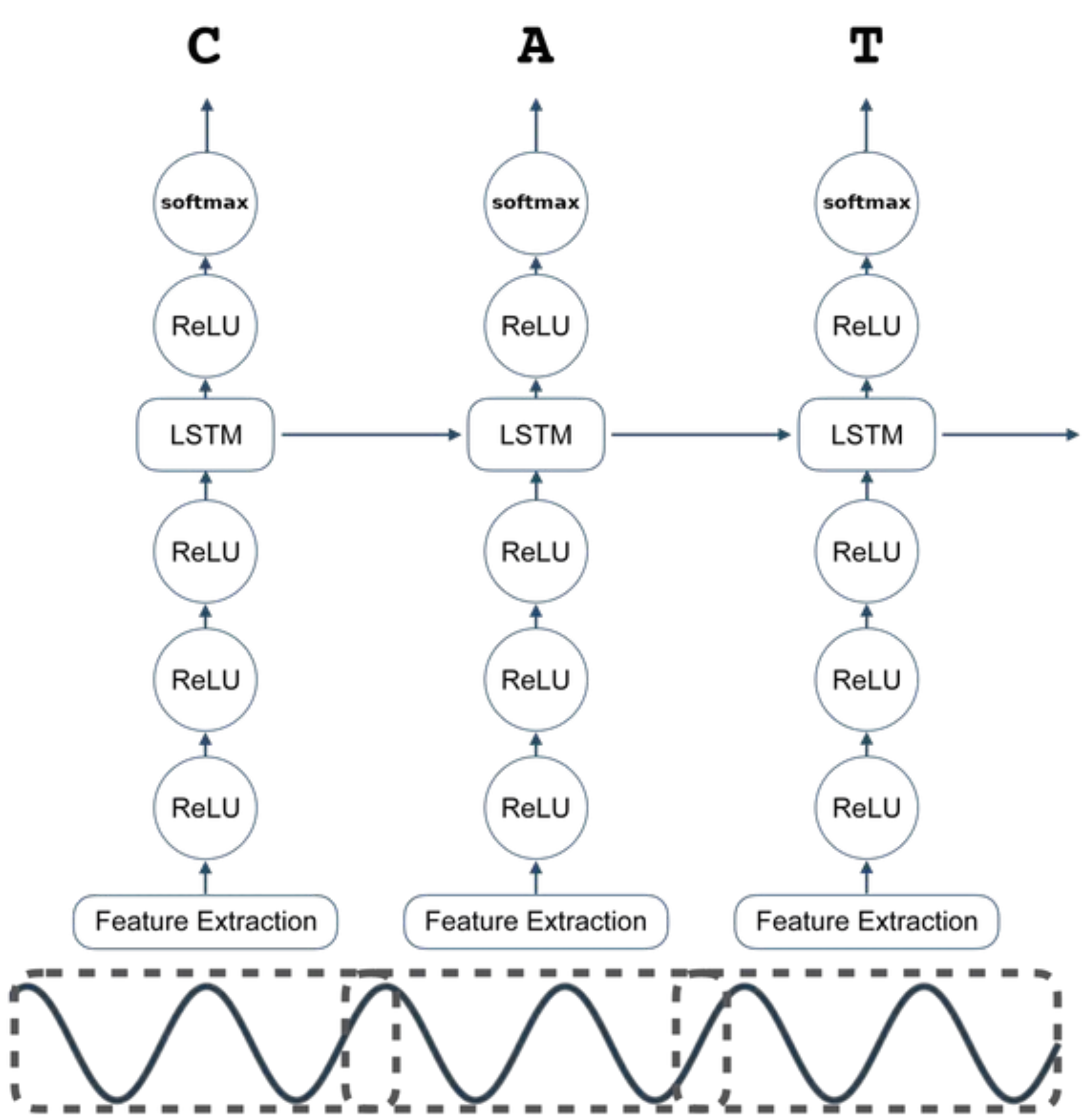
	Gender	Duration (hrs)
Transcribed	—	20
Untranscribed	Women	1.4
	Men	4.6
	—	129
Total		**155**

Table 1: Released radio dataset

	Tokens	Types	Hours
Training	900,608	135,647	107.1
Validation	99,839	27,939	11.8
Testing	14,117	5,110	1.8
Total	—	—	120.7

Table 2: Dataset used for Training

Architecture



Experiments and Results

Dataset	WER (%)
Common Voice	33
Radio	47

Table 3: WER on CV and Radio dataset.

Gender	WER	Duration (mins)
Women	70.6%	14
Men	53.5%	14

Table 4: Performance on a held-out test

Boost Value	Transcript	Verdict
-1000	... abantu balina okwegendereza **ekifo** tulina gugaawulira ...	false negative
-600	... abantu balina okwegendereza **ekifo** tulina gugaawulira ...	false negative
-200	... abantu balina okwegendereza **ekifo** tulina gugaawulira ...	false negative
0	... abantu balina okwegendereza **ekifo** tulina gugaawulira ...	false negative
+200	... abantu balina okwegendereza **ekifuba** e u o i na gugaawulira ...	true positive
+600	... abantu balina okwegendereza **ekifuba** e u o i na gugaawulira ...	true positive
+1000	... abantu balina okwegendereza **ekifuba** e u o i na gugaawulira ...	true positive

Table 5: Keyword was mentioned in the audio but the ASR had failed to transcribe it

Keyword	ASR		HTWD-B	
	TP	FN	TP	FN
"covid"	71	12	71	12
"ekirwadde"	11	5	14	2
"kolona"	3	3	6	0
"ssennyiga"	5	0	5	0
"ekifuba"	1	1	2	0

Table 6: TP and FN on COVID-19 test set

Metric	ASR	HTWD-B
True Positives	91	98
False Positives	0	6
False Negatives	21	14
Precision	1	0.99
Recall	0.81	0.89
Fscore	0.89	0.94

Table 7: Fscore results for ASR and HTWD-B

Conclusion

We present the Makerere Radio Speech Corpus Luganda radio corpus and Luganda ASR for radio monitoring. We show how we utilized transfer learning to fine tune a Kinyarwanda model on Luganda Common Voice and radio data. We present results on a publicly available Luganda Common Voice and on a radio dataset. We evaluate the model's performance on a held-out test-set of COVID-19 keywords to obtain Fscore of 0.94. We highlight the importance of gender consideration in ASR models by evaluating our model on women's and men's voices. We believe that this work has the potential to benefit many researchers working on radio monitoring work in sub-Saharan Africa.

References

- Menon, R., Saeb, A., Cameron, H., Kibira, W., Quinn, J., and Niesler, T. (2017). Radio-browsing for developmental monitoring in uganda. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5795–5799. IEEE.
- Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., and Weber, G. (2019). Common voice: A massively-multilingual speech corpus. arXiv preprint arXiv:1912.06670.