Modeling Bias in Automatic Speech Recognition

Camille Harris
Georgia Institute of Technology
charris320@gatech.edu

Chijioke Mgbahurike
Stanford University
cmgbahur@stanford.edu

Diyi Yang
Stanford University
diyiy@stanford.edu

Introduction

Background

Gender and dialect bias has impacts accuracy of automatic speech recognition. Minorities and women who use speech technologies may struggle with inaccuracies.

Our Contributions

- ★ Dataset of audio data labeled for dialect and gender, including underrepresented dialects
- ★ Benchmark analysis of ASR systems and their performance on various demographic groups
- ★ Analysis of the impact of fine-tuning methods on performance of ASR models on marginalized groups

Methods

Dialect-Centered Data Collection and Annotation

- ★ Identity related keywords used to identify podcasts from the Spotify Podcast Dataset
- ★ Dialect speakers transcribe and timestamp podcast data, and record metadata about the podcast

Benchmarking

- ★ Analysis of performance of ASR models across dialect, gender, and dialect-gender combined categories
- ★ Future analysis will include impact of fine-tuning methods

Results

Benchmarking Base Models with Word Error Rate

Whisper	HuBERT	Wav2Vec2
0.28	0.296	0.389

	Men	Women	Men and Women
Whisper	0.316	0.259	0.26
HuBERT	0.355	0.245	0.323
Wav2Vec2	0.467	0.323	0.416

	AAVE and SAE	AAVE only	Chicano English	Spanglish	Chicano English and Spanish	SAE + other dialect	SAE only
Whisper	0.2702	0.374	0.441	0.405	0.494	0.263	0.247
HuBERT	0.297	0.530	0.421	0.437	0.608	0.265	0.252
Wav2Vec2	0.395	0.6451	0.526	0.550	0.712	0.354	0.333

	AAVE		Chicano English		Spanglish		SAE	
	Men	Women	Men	Women	Men	Women	Men	Women
Whisper	0.291	0.232	0.458	0.391	0.408	0.342	0.215	0.26
HuBERT	0.323	0.246	0.522	0.134	0.523	0.278	0.244	0.265
Wav2Vec2	0.422	0.358	0.633	0.189	0.641	0.36	0.311	0.330

Conclusion

Conclusion

We find generally better performance for SAE speakers compared to minority dialects

Minority dialect speech performs better when combined with code-switching to SAE

Limitations

- ★ Dialect overlap and code-switching
- ★ Imbalanced categories in the dataset
- ★ Code-switching imbalance by gender

Future Work

- ★ Updated iteration of the dataset
- ★ Analysis of the impact of Vanilla and LoRA fine-tuning approaches

Acknowledgements

Camille Harris is generously supported by the Ford Fellowship, the GEM Fellowship, and the Amazon Consumer Robotics Grant

Contact info

Email: charris320@gatech.edu
Twitter/X: @CamilleAHarris
Website: camille2019.github.io