Exploring Subgroup Performance in End-to-End Speech Models

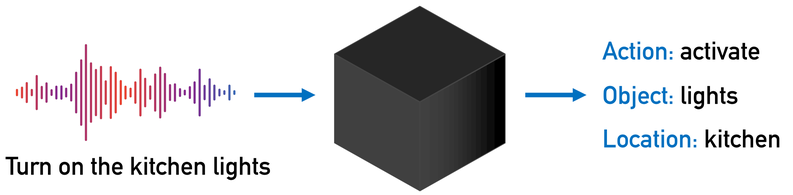
Alkis Koudounas, Eliana Pastor, Giuseppe Attanasio, Vittorio Mazzia, Manuel Giollo, Thomas Gueudre, Luca Cagliero, Luca de Alfaro, Elena Baralis, Daniele Amberti

Politecnico di Torino | Amazon Alexa AI | Bocconi University | University of California

48th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023)

E2E SLU Systems

End-to-end (E2E) spoken language understanding (SLU) models achieve high performance but are **complex black-box processes**.



Investigating problematic data subgroups is crucial for understanding, debugging and ensuring model fairness.

Subgroups: defined via a set of metadata, representing user information, recording and speech conditions → e.g., gender=female, age=22-40

Our Analysis

Individual Model Analysis: study the subgroup-level model performance of a speech E2E model.
Divergence as a measure of anomalous behavior of a data subgroup S w.r.t. overall dataset D for a function f.

$$\Delta(S) = f(S) - f(D)$$

Subgroup-Level Model Comparison: compare different models, identifying the subgroups where performance improves or suffers when changing model.
Subgroup gain as the difference in performance of two models M_1, M_2 on a specific subgroup S for f.

$$gain_f(S, M_1, M_2) = f(S, M_2) - f(S, M_1)$$

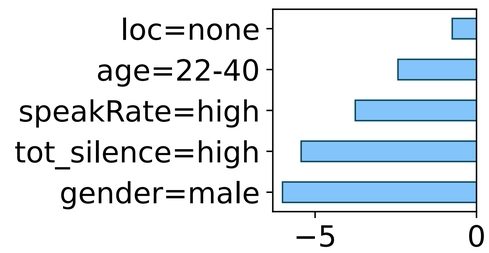
Individual Model Analysis

Fluent Speech Commands (FSC) dataset
wav2vec 2.0 large model

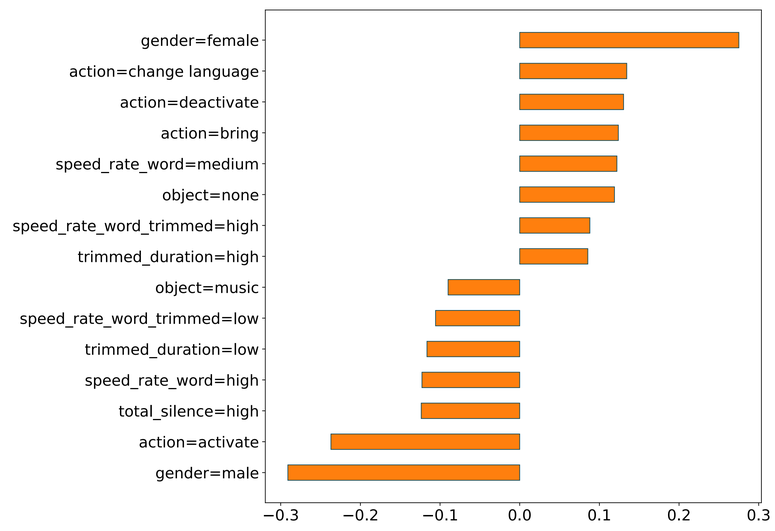
Most divergent subgroups

Subgroup	*Sup*	*acc*	Δ_{acc}	t
{*age=22-40, gender=male, loc=none, speakRate=high, tot_silence=high*}	0.03	74.79	-18.38	4.7
{*action=increase, gender=male, speakRate=high*}	0.03	74.81	-18.36	4.9

Shapley values: contribution of each term within the subgroup

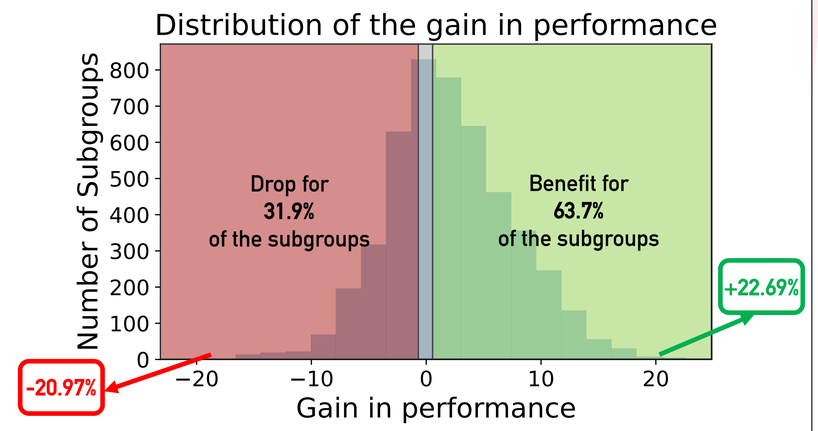


Global Shapley values: global contribution of each term across all subgroups

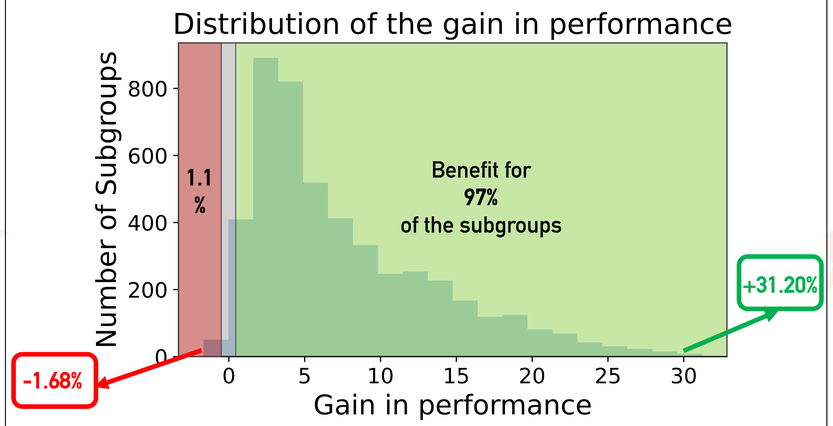


Subgroup-Level Model Comparison

wav2vec 2.0 base and large models



wav2vec 2.0 base and HuBERT base models



Scan for further info!