

PRIORITIZING DATA ACQUISITION FOR END-TO-END SPEECH MODEL IMPROVEMENT

ICASSP 2024



WHO?

Bocconi
UC SANTA CRUZ

Alkis Koudounas, Eliana Pastor, Giuseppe Attanasio, Luca de Alfaro, Elena Baralis
Politecnico di Torino, Italy; Bocconi University, Italy; University of California, USA

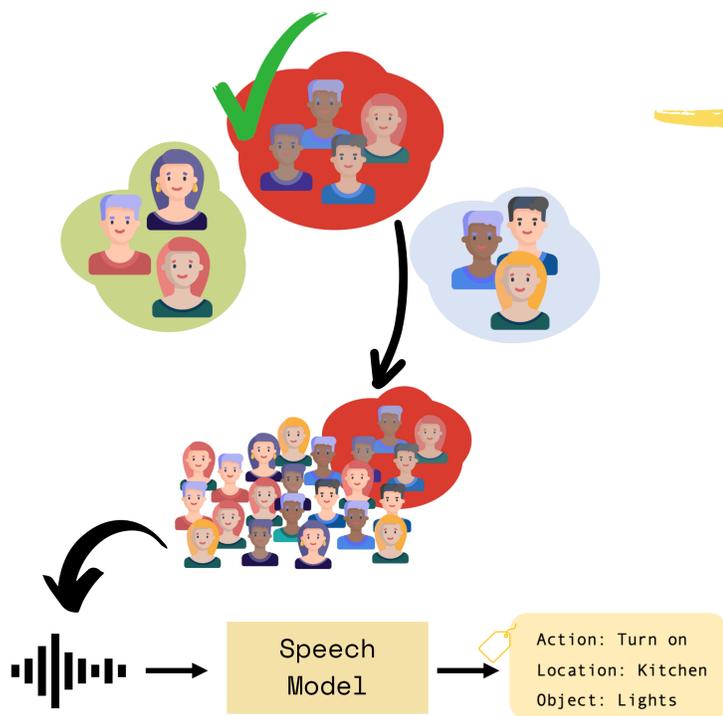
WHY?

1. Can we enhance model performance, overall and across subgroups?
2. Is the mantra “The more data, the better” true?

WHAT?

Novel *divergence-aware* acquisition guided by interpretable problematic subgroups of data

HOW?



1. Automatic detection of *interpretable* and *underperforming* subgroups through the concept of *divergence*:

$$\Delta^-(S) = f(S) - f(D) < 0$$

2. Acquisition of *only* the samples belonging to k most *underperforming* subgroups

3. Speech model re-training

YES, BUT... THE RESULTS?

FSC

K	Approach	#samples	Accuracy	F1 Macro	Δ_{max}^-	Δ_{avg-10}^-	Δ_{avg-20}^-	Δ_{avg-50}^-	$ \Delta_{avg-all} $
-	original	-	91.58 ± 0.08	86.34 ± 0.13	-70.09 ± 0.26	-70.09 ± 0.26	-65.73 ± 0.49	-53.31 ± 0.19	1.06 ± 0.07
	random	406	94.26 ± 0.27	91.17 ± 0.86	-54.26 ± 1.14	-53.93 ± 1.17	-53.24 ± 1.12	-52.37 ± 0.55	0.86 ± 0.06
	random	226	92.56 ± 0.44	90.25 ± 0.60	-52.20 ± 2.57	-51.11 ± 2.19	-46.61 ± 1.34	-43.98 ± 0.68	0.97 ± 0.02
2	clustering	406	92.94 ± 0.07	90.82 ± 1.19	-51.81 ± 0.86	-51.22 ± 0.92	-49.99 ± 0.10	-48.52 ± 0.11	1.24 ± 0.09
	clustering	226	89.77 ± 0.88	87.02 ± 0.15	-47.37 ± 0.42	-47.34 ± 0.42	-47.23 ± 0.43	-46.75 ± 0.91	0.94 ± 0.04
	<i>ours</i>	226	96.55 ± 0.08	94.71 ± 0.12	-40.60 ± 0.35	-40.28 ± 0.36	-38.08 ± 0.36	-32.72 ± 0.28	0.81 ± 0.03
-	all data	4606	93.42 ± 0.17	93.11 ± 0.17	-53.18 ± 0.15	-50.89 ± 0.09	-45.61 ± 0.14	-40.37 ± 0.16	0.37 ± 0.01

We obtain the **best** performance **overall and over subgroups**

Evaluation on **Intent Classification** in English (FSC) and Italian (ITALIC)

WHERE?

