



To remember or Not to Remember?

Continual Learning from Adult to Children for Low-resource Automatic Speech Recognition

Maria Garofalaki, Line H Clemmensen¹, Sneha Das¹

Motivation

- State-of-the-art models need large datasets during training, for optimal performance.
- Inadequate development of ASR models for low-resource languages and atypical populations.
- Solution: Pre-training+ fine-tuning?
- Leads to catastrophic forgetting

Objective

ASR for adult-child conversations with minimal labelled data

Contributions

- Mitigate data-constraints: explore augmentation
- Systemic investigation of weight-freezing for continual learning from adults to children ASR.
- Elastic weight consolidation: retain important information from adults ASR + learn new aspects of children's ASR.
- Present results on the unseen clinical dataset with naturalistic conversations.

Results

1. Impact of data-augmentation
2. Weight freezing (vs) Elastic weight Consolidation (EWC) on:
 - a. Adult and children
 - b. Clinical data
3. Final models

	Dataset	Duration (h:m)	Number	Group	Type
1.	Danpass	01:21	27	Adults	Fine-tuning
2.	Common-voice	02:20	99	Adults	Fine-tune
3.	Plunkett-1	00:25	1	Child	Fine-tune + Eval.
4.	Plunkett-2	00:04	1	Child	Eval.
5.	Alvenir	05:00	50	Adults	Eval.
6.	Clinical-data	06:00	12	Child	Eval.

1	Dataset	Augmentation-type			
		FT	A1	A2	A1+A2
1.	Alvenir	.27 ± .22	.62 ± .31	.3 ± .23	.62 ± .31
2.	Plunkett	1 ± .2	.98 ± .15	1 ± .2	.97 ± .15
3.	Clinical	.75 ± .002	.9 ± .02	.74 ± .07	.87 ± .04

1. Department of Applied Mathematics and Computer Science, Technical University of Denmark

Methodology

- Pre-trained model: xls-r-300m-danish-nst-cv9
- Data Augmentation:
 - Child-like characteristics to training speech samples
 - Noise augmentation for robustness
- Continual Learning
 - Weight freezing
 - Elastic weight consolidation

