DOMAIN ADAPTATION FOR SPEECH RECOGNITION THROUGH SEMI-SUPERVISED LEARNING

Oliver Bentham   Dr. Stephen LaRocca

Introduction

Semi-Supervised Learning is a type of machine learning that makes use of unlabeled data for training to improve accuracy, and has been shown to work in Automatic Speech Recognition[1].

Domain Adaptation is the process of modifying a model trained on data from a specific source domain to improve performance on a target domain.

Our goal is to find the smallest quantity of out-of-domain speech data to train a speech recognition system whose predictions on unlabeled in-domain data improve accuracy during semi-supervised learning.

Motivation

The Multilingual Computing and Analytics Branch at ARL has speech-to-speech devices deployed with soldiers in noisy, conversational environments.

Most speech data available in low-resource languages is clean audio from parliamentary proceedings or TV news

A method to train low-resource, robust ASR models for everyday speech would benefit most ASR applications

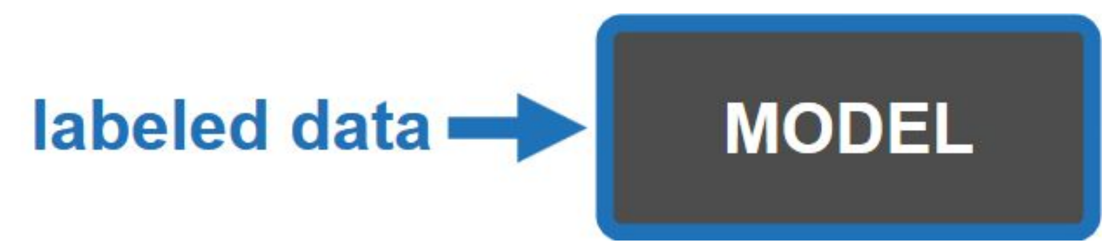
Data

We use two very common English ASR corpora, Wall Street Journal (WSJ) and Switchboard (SWBD)

speech corpus	domain		labeled/ transcribed
Wall Street Journal	out-of-domain	edited news clean	yes
Switchboard	in-domain	conversational phone noisy	no

Approach

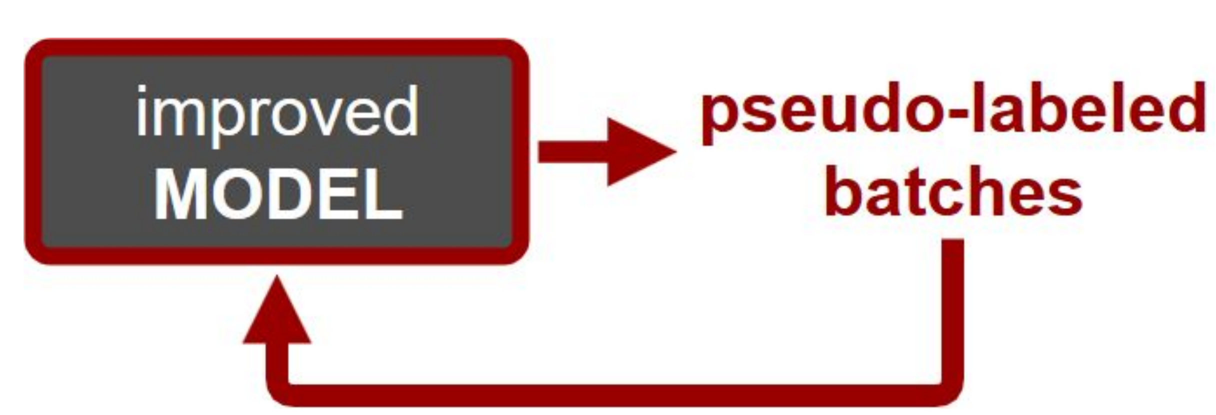
1. **Prepare the data.** Split the out-of-domain **WSJ** training data into subsets
2. **Train a model.** Using Kaldi, train an LF-MMI chain model[2] on each of the **WSJ** training subsets to get split1, split2, split3, etc.



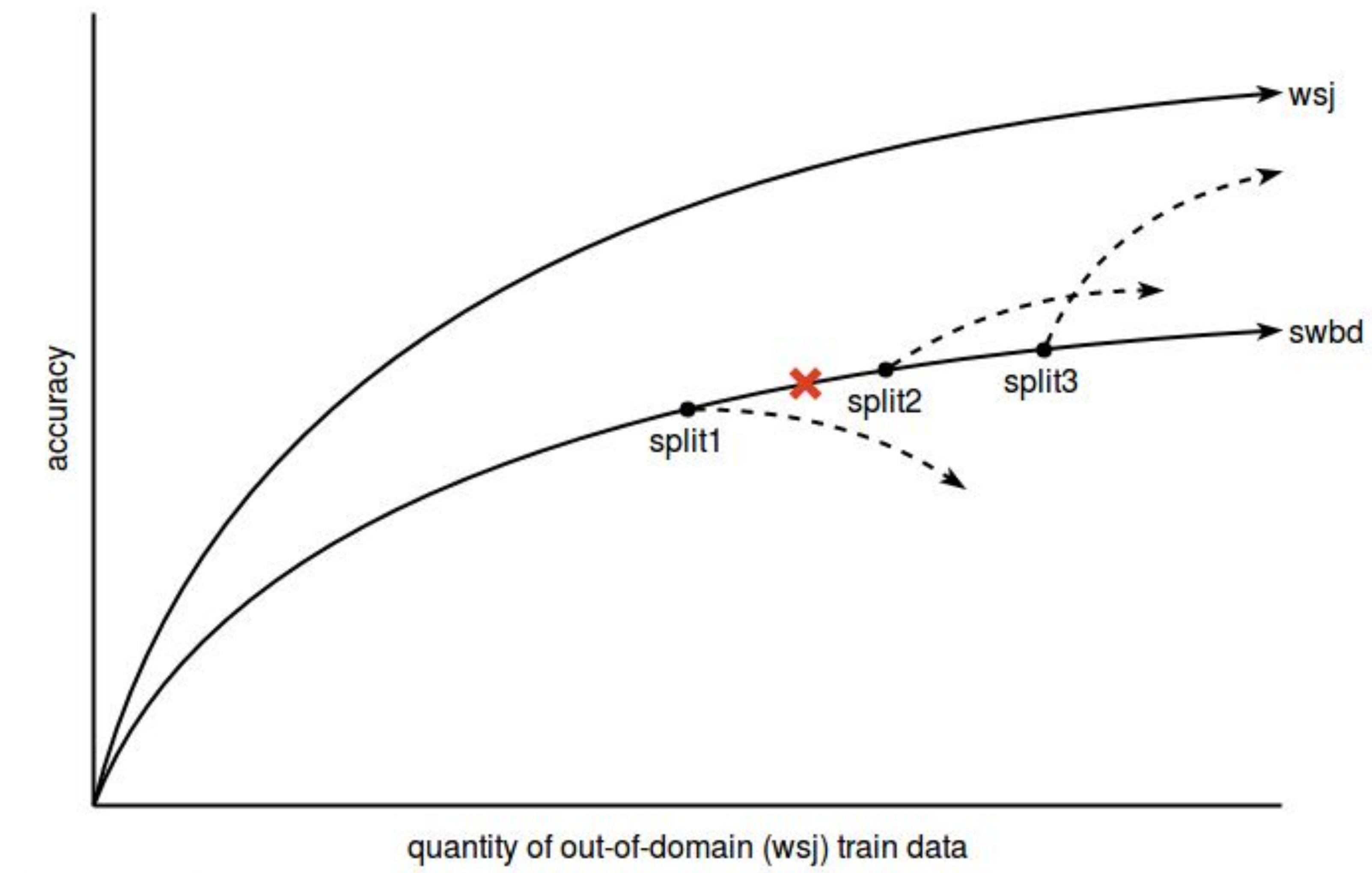
3. **Assign pseudo-labels.** Use each split's model to assign pseudo-labels to a batch of the unlabeled **SWBD** data

4. **Train improved model.** Continue to train the model on the pseudo-labeled **SWBD** batch to get an improved model

5. **Repeat.** Return to step 3 with improved model and incorporate another **SWBD** batch



Hypothesized Results



Results

Baseline Word Error Rates for splits prior to semi-supervised learning

wsj training utterances	wsj test set	swbd test set
5k	11.67	94.76
10k	9.03	86.95
15k	7.60	85.98
20k	7.23	83.74
25k	7.20	**72.96**
30k	**7.03**	79.91

Future Work

1. Test with different languages and different domains
2. Experiment with different language models
3. Investigate different neural architectures

References

[1] Manohar, V., Hadian, H., Povey, D., & Khudanpur, S. (2018). Semi-Supervised Training of Acoustic Models Using Lattice-Free MMI. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp.2018.8462331

[2] Povey, D., Peddinti, V., Galvez, D., Ghahremani, P., Manohar, V., Na, X., . . . Khudanpur, S. (2016). Purely Sequence-Trained Neural Networks for ASR Based on Lattice-Free MMI. Interspeech 2016. doi:10.21437/interspeech.2016-595