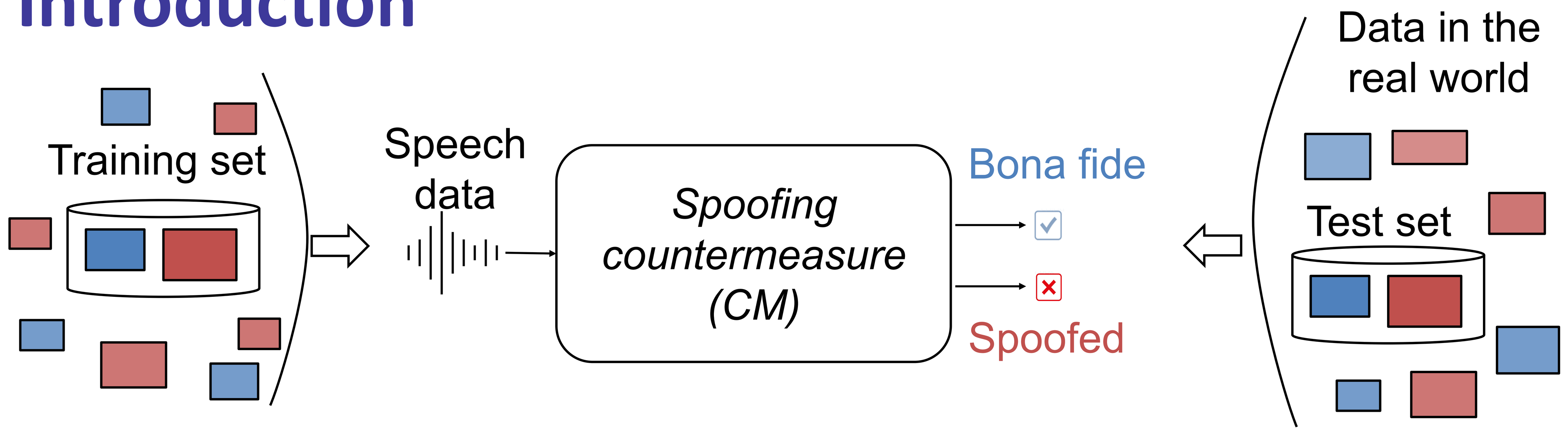
3-1-15-SLR
SLT2022

Investigating Active-learning-based Training Data Selection for Speech Spoofing Countermeasure

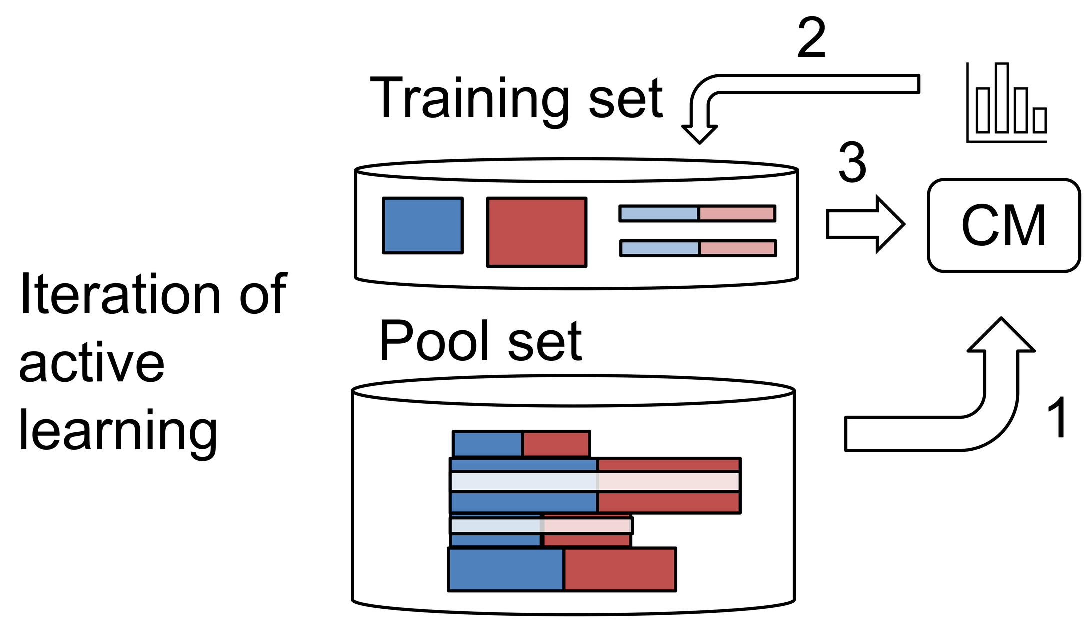
Xin Wang, Junichi Yamagishi, National Institute of Informatics

Introduction



- A common practice: train a CM using a training set; test its performance on a test set; both sets are from a standard database.
  - The trained CM may not well generalize [1,2,3].
  - More training data?
- In real scenarios where abundant data is available, **can the CM *automatically select useful data* during training**?

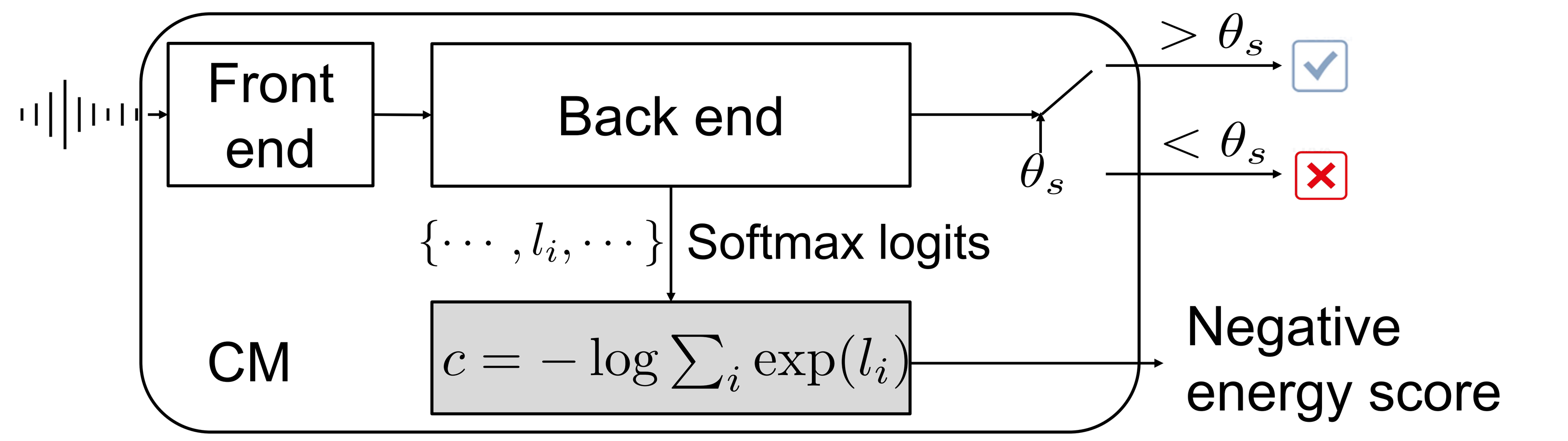
Training by active learning[4] (AL)



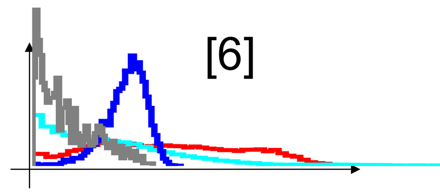
Pre-train CM on a seed training set
1. Measure the *usefulness* of pool set data
2. Select and add useful data to the training set
3. Fine-tune CM, go to step 1

How to estimate the *usefulness* of the data? (Sec.2.1)

- One metric: negative energy score [5]. It measures whether the data is out of training data distribution.



Negative energy scores of In-dist. spoofed, in-dist. bona fide, Out-of-dist. spoofed, out of dist. bona fide data



Alternative AL approach: actively remove *useless* data from the pool and randomly select (Sec.2.2)

Experiments & findings

CM configurations (following our previous work [3])

- Front end: wav2vec2.0 XLSR-53, fine-tuned with the back end
- Back end: global average pooling, a linear layer, softmax

Datasets

- Seed training set: ASVspoof 2019 logical access training set
- Pool sets were created from databases with diverse attacking methods (att.), languages (lang.), and speakers (spk.).

(a) Base data subsets

ID	Source	#. Trial	Dur. (h)	Lang.	#. Spr	#. Att
①	ASVspoof2019 LA trn.	2,580 / 22,800	24.0	En	20	6
②	FMFCC-A, trn.	4,000 / 6,000	5.5	Zh	77	5
③	ESPNet on LibriTTS	736 / 4,275	8.0	En	>80	6
④	ESPNet on LJSpeech	200 / 1,800	3.6	En	1	9
⑤	BC 2019	75 / 5,925	15.5	Zh	1	25
⑥	VoxCeleb1	6,000 / 0	13.6	Mul.	>1 k	0

(b) AL seed and pool sets created from base data subsets

	Base data	#. Trial	Dur. (h)	Lang.	#. Spr	#. Att.
Seed set	①	2,580 / 22,800	24.0	En	20	6
Pool set A	②+③	4,736 / 10,275	13.5	En, Zh	>150	11
Pool set B	②+③+④+⑤+⑥	11,011 / 18,000	46.3	Mul.	>1.1 k	45

- Test sets: ASVspoof 2019, 2021 (LA & DF), and WaveFake [7]

[1] Paul, et al A Case Study with ASVspoof 2015 and BTAS 2016 Corpora. In Proc. ICASSP, 2047–2051, 2017
[2] Das, et al. Assessing the Scope of Generalized Countermeasures for Anti-Spoofing. In Proc. ICASSP, 6589–6593. 2020.
[3] Wang, X., et al. Investigating Self-Supervised Front Ends for Speech Spoofing Countermeasures. In Odyssey, 2022
[4] Settles, B. Active Learning Literature Survey. 2009.
[5] Liu, W. et. al., Energy-Based Out-of-Distribution Detection. In Proc. NIPS, 33:21464–21475. 2020.
[6] Wang, X., et al. Estimating the Confidence of Speech Spoofing Countermeasure. In Proc. ICASSP, 6372–6376. 2022.
[7] Frank, et. al., WaveFake: A Data Set to Facilitate Audio DeepFake Detection. In Proc. NeurIPS Datasets and Benchmarks 2021.
[8] Chandra, et al. On Initial Pools for Deep Active Learning. In NeurIPS 2020 Workshop. 2021.

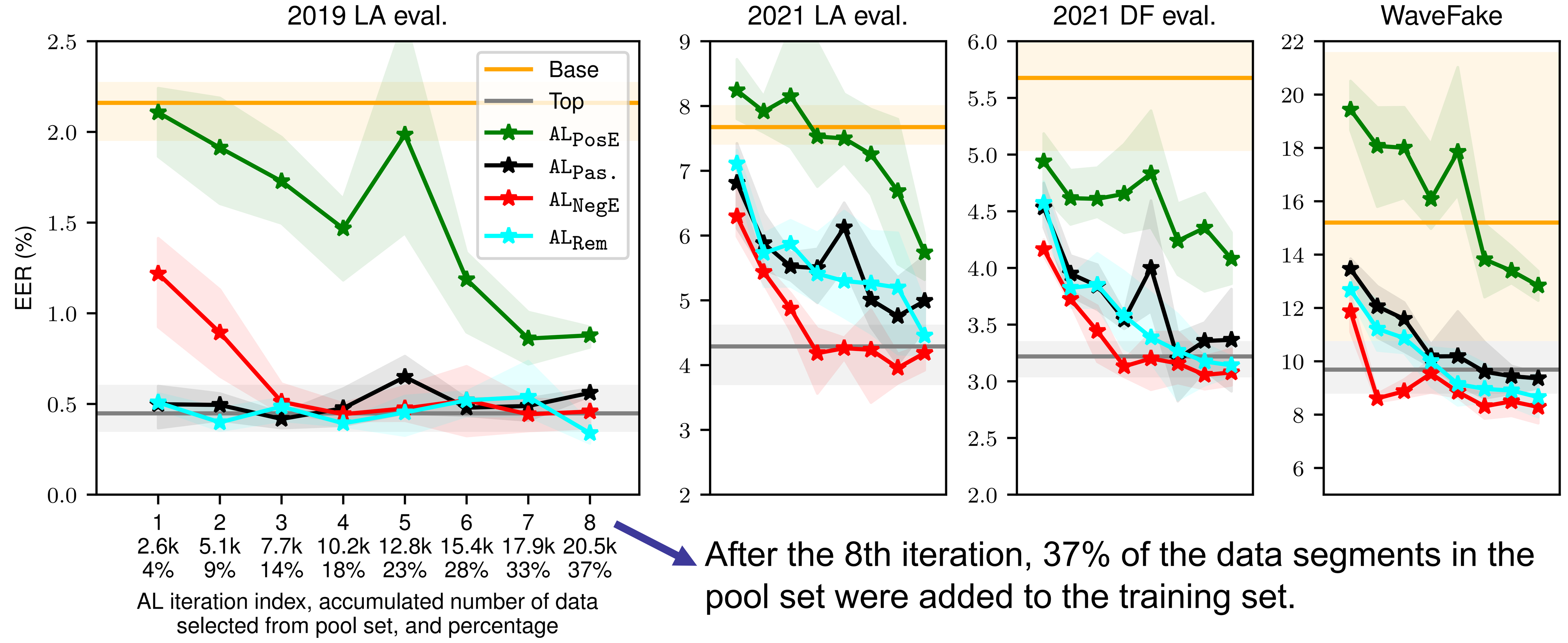
Experimental systems:

Conventional training methods
- Base: trained w/ seed set
- Top: trained w/ seed & pool set B

AL methods using pool set B
- AL_{NegE}: negative energy score
- AL_{PosE}: similar to AL_{NegE} but with the energy score value reversed
- AL_{Pas}: random energy score
- AL_{Rem}: active data removing w/ negative energy score

Equal error rates (EER %) on four test sets

Each line shows max, average, and min EER values after running the experiment three times

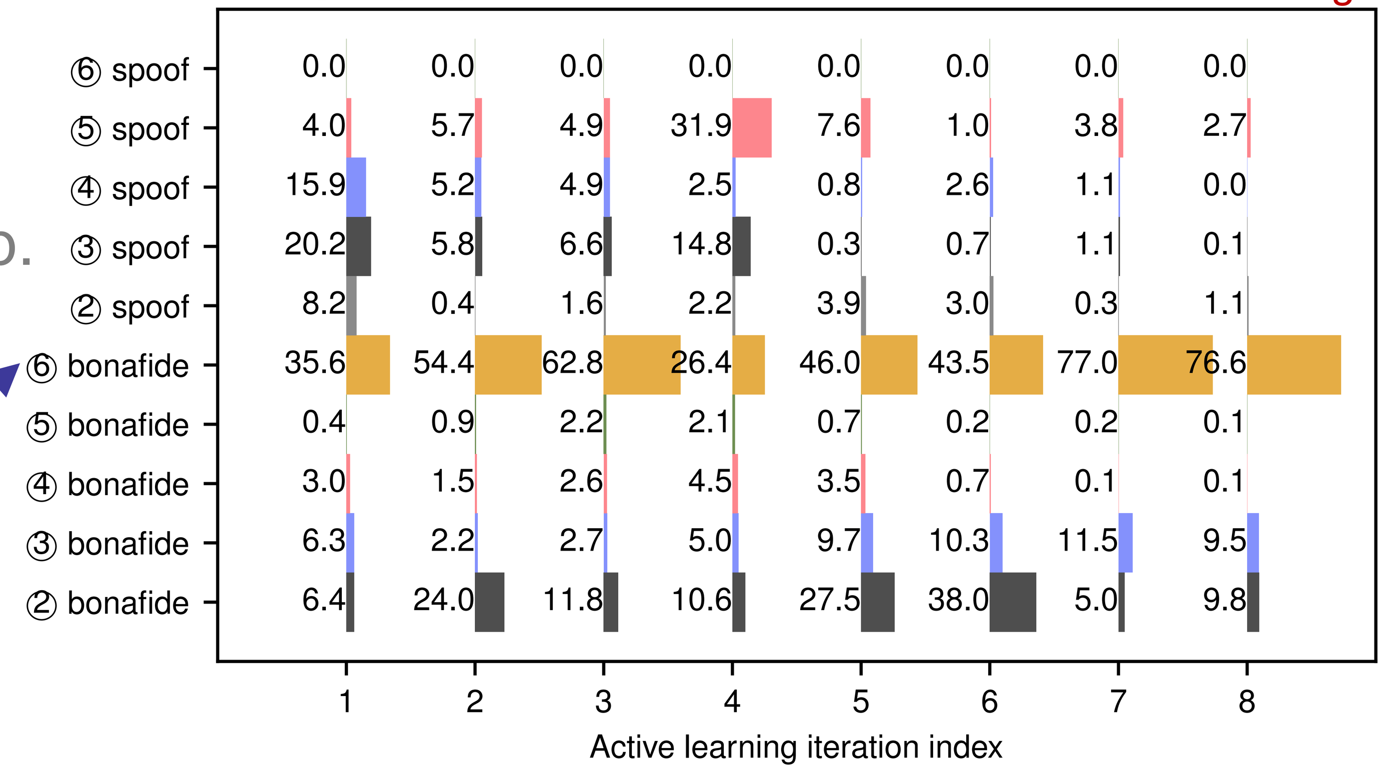


After the 8th iteration, 37% of the data segments in the pool set were added to the training set.

Highlighted findings:

- Can the CM *automatically select useful data* and train itself? **Yes**
  - Good AL methods (e.g., AL_{NegE}) are more data efficient than Top.
  - However, do NOT select useless data (AL_{PosE}).
  - Simple random data selection (AL_{Rem}) is good [8].
- What kind of pool data was found to be useful? VoxCeleb1
- Other findings: do we need a good pool set? Yes, the pool set should have diverse data (see sec.3.3).

Percentage (%) of pool set data selected by AL_{NegE}



Inter-University Research Institute Corporation / Research Organization of Information and Systems
National Institute of Informatics

Contact: wangxin@nii.ac.jp, Code: https://github.com/nii-yamagishilab/project-NN-Pytorch-scripts project/08-asvspoof-activelearn
Partially supported by JST CREAST grant JPMJCR18A6, JPMJCR20D3, JSPS KAKEN 21K17775, 21H04906, 18H04112, and Google AI for Japan program.