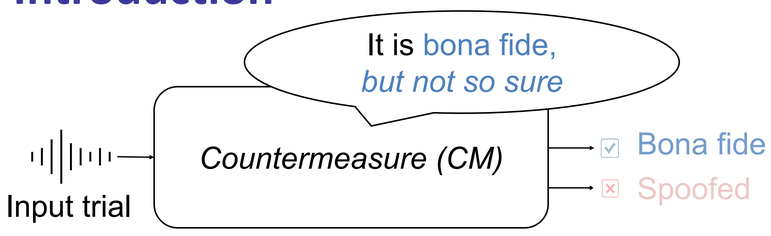
Estimating Confidence of Speech Spoofing Countermeasure

Xin Wang, Junichi Yamagishi, National Institute of Informatics, Japan

ICASSP 2022
Paper ID 2903

Introduction



Existing CMs usually do not produce confidence.

CM confidence may be informative:

- abstention: avoid risky decisions for low-confidence trials
- active learning: annotate them and add to training data [1]

How to estimate confidence scores? Any benefit?

Methods to estimate confidence *c*

Let $\{P_0, P_1\}$ be prob. of {*spoofed* and *bona fide*}.

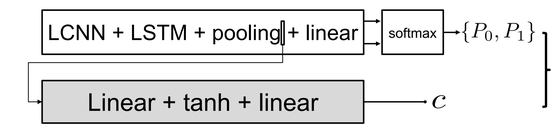
Let $\{l_0, l_1\}$ be logits of softmax output layer.

I Max-prob. estimator[2]: $c = \max\{P_0, P_1\}$

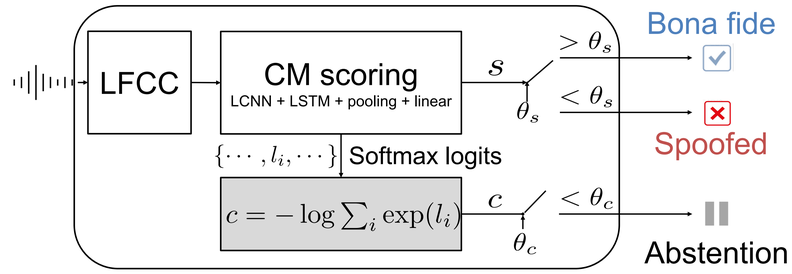
II Negative energy[3]: $c = -\log \sum_i \exp(l_i)$

III Min negative Mahalanobis dis. [4]

IV Jointly trained DNN conf. estimator[5]:



V Independent DNN conf. estimator (require labels)



	Score range	Trainable?
Max prob.	$c \in [0.5, 1)$	No
Energy score	$c \in \mathbb{R}$	No
Neg. M-dist.	$c \in (-\infty, 0)$	Trained after CM
Conf. branch	$c \in (0, 1)$	Jointly trained with CM
Supervised	$c \in (0, 1)$	Separately trained

Experiments

Model config

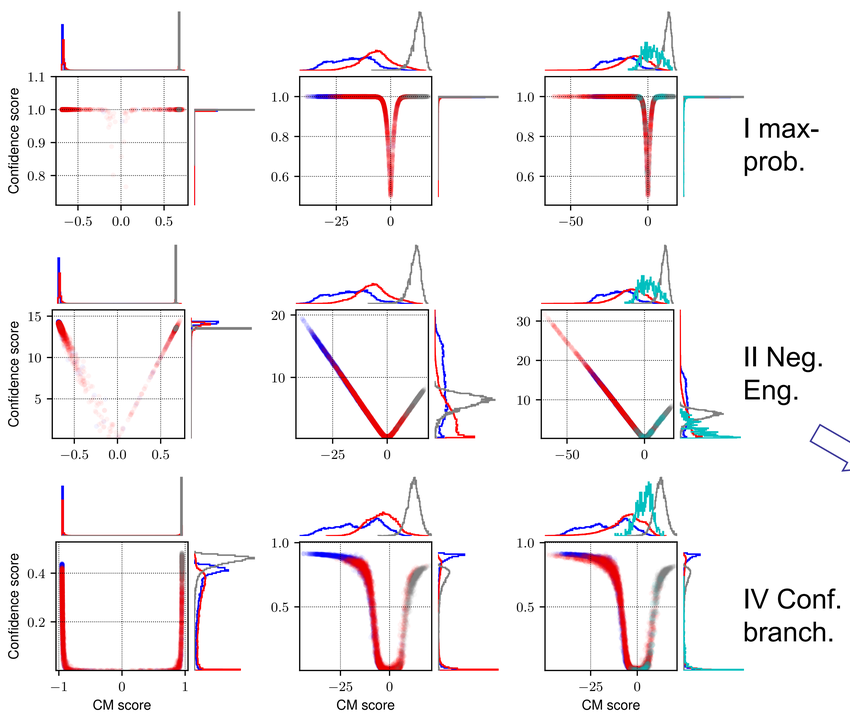
- LFCC, LCNN + LSTM + pooling + linear[6]
- Either additive margin (AM) or vanilla plain softmax

Database & protocols

- *unknown* trials should receive low conf., i.e., they do not match with training data in terms of channel, language, …
- *known* trials should receive high conf..
- Two test sets to simulate different testing scenarios.
- Let's focus only on T1 training set, which is is the standard ASVspoof 2019 LA training set.

		Known trials	*Unknown* trials
Train set	T1	LA trn. (2,580 / 22,800)	-
	T2	LA trn. (2,580 / 22,800)	ESPNet (250 / 2,000)
	T3	LA trn. (2,580 / 22,800)	BC19 (100 / 7,625)
Test set	E1	LA test kn. (7,355 / 19,656)	LA test unk. (0 / 44,226)
	E2	LA test kn. (7,355 / 19,656)	VCC (770 / 49,467)

[1] Settles, B. Active Learning Literature Survey. 2009.
[2] Hendrycks D., A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In *ICLR*. 2017.
[3] Liu, W., et al.. Energy-Based Out-of-Distribution Detection. In *Proc. NIPS*, 33:21464–21475. 2020.
[4] Lee, K., et. al. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. in *NIPS*, 2018.
[5] DeVries, T. & Taylor, G. W. Learning confidence for out-of-distribution detection in neural networks. arXiv1802.04865, 2018
[6] Wang, X. et. al. A comparative study on recent neural spoofing countermeasures for synthetic speech detection. in *Interspeech* 2021



Unknown spoofed and Unknown bona fide should have lower conf. scores.

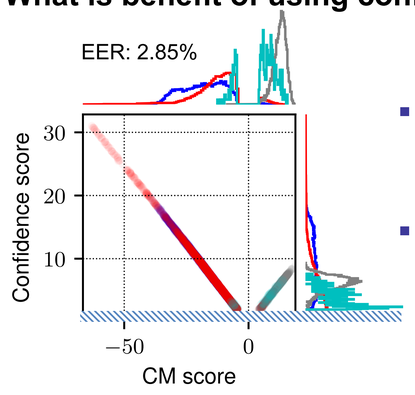
Known spoofed and Known bona fide should have higher conf. scores.

Messages

Can conf. scores be estimated?

- x max-prob. and a few methods
- ✓ Promising: Neg. energy and conf. branch are
- With AM-softmax, no, CM scores follow bi-nomial dist. – the CM is over-confident
- With plain softmax, yes

What is benefit of using confidence scores?



- Avoid making decisions on trials w/ low conf. scores
- Select useful training data w/ low conf. (see arxiv:)

Limitation: unknown unknowns are difficult to detect, for example, unknown bona fide data with low CM scores

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

Partially supported by JST CREAST grant JPMJCR18A6, JPMJCR20D3, JSPS KAKEN 21K17775, 21H04906, 18H04112, and Google AI for Japan program.