One-Class Learning Towards Synthetic Voice Spoofing Detection

You Zhang, Fei Jiang, Zhiyao Duan
University of Rochester

ABSTRACT
Automatic Speaker Verification (ASV) systems are vulnerable to text-to-speech (TTS), and voice conversion (VC) attacks. Voice anti-spoofing is developed to improve the reliability of speaker verification systems against such spoofing attacks. The fast development of speech synthesis are posing increasingly more threat.

Main issue of voice anti-spoofing systems:
- Generalization to unseen synthetic attacks

Proposed solution:
- One-Class Learning

Results:
- EER 2.19%, outperforming all single systems

Keywords:
Voice Anti-spoofing, One-Class Learning, Generalization Ability, Feature Learning, Speaker Verification, Voice Biometrics

BACKGROUND
Automatic Speaker Verification (ASV)
Verify the identity of a speaker

Logical Access (LA) Spoofing Attacks
- Text-to-speech (TTS)
- Convert written text into audio with speech synthesis
- Voice Conversion (VC)
- Convert speech from source to a target speaker

Voice Anti-spoofing / Spoofing Countermeasure (CM)
Detect spoofing attacks

METHOD
One-Class Learning
The distribution mismatch between training and test for the spoofing attacks class, makes the problem a good fit for one-class classification [1].

We propose a loss function called one-class Softmax (OC-Softmax) to learn a feature space in which the bona fide speech embeddings have a compact boundary while spoofing data are kept away from the bona fide data by a certain margin.

The proposed OC-Softmax can be formulated as:

\[ L_{OC-Softmax} = \frac{1}{N} \sum_{i=1}^{N} \log \left( 1 + \exp \left( \frac{c_{i} - \alpha}{\sigma} \right) \right). \]

Preliminary: Binary classification loss functions
Vanilla Softmax

\[ L_{S} = -\frac{1}{N} \sum_{i=1}^{N} \log \left( \frac{e^{x_{i,j}}}{\sum_{j=1}^{M} e^{x_{i,j}}} \right) \]

Additive Margin Softmax [2]:

\[ L_{AMS} = -\frac{1}{N} \sum_{i=1}^{N} \log \left( \frac{e^{w_{i,j}^{\alpha} + \delta}}{\sum_{j=1}^{M} e^{w_{i,j}^{\alpha} + \delta}} \right) \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \log \left( 1 + \exp \left( \frac{\alpha}{\sigma} \right) \right). \]

RESULTS
Dataset: ASVspoof 2019 LA

We evaluate the performance with equal error rate (EER) and minimum tandem detection cost function (i-DCF).

Comparing with binary classification loss functions:

Our proposed OC-Softmax achieves the best results.

Feature embedding visualization:

CONCLUSIONS
- One-class learning aims to compact the target class representation in the embedding space, set a tight classification boundary around it and push away non-target.
- One-class learning could improve the generalization ability of anti-spoofing system against unknown spoofing attacks.
- The proposed system trained with OC-Softmax outperforms all existing single systems.

ACKNOWLEDGMENTS

REFERENCES

CITATION

FOLLOW-UP WORKS