ECE 209AS - Project Midterm Presentation

Automatic Speech Verification - Spoof Detection

Sidarth Srinivasan (005629203)
Nithin Varma (005851269)

May 11, 2022

Project Goals

- The Automatic Speaker Verification (ASV) system ideally aims to verify the identity and authenticity of a target user given an audio sample.
- However, these ASV systems are vulnerable to spoofing attacks of the following kind:
  - Impersonation attacks
  - Replay
  - Speech Synthesis
  - Voice Conversion
- Application: User Authentication (eg. banks, call centres, smart phones etc.)
- Goal of the project is to develop a Countermeasure (CM) system to complement the ASV system to verify the authenticity (original/fake) of a given audio sample.

Specific Aims - Countermeasure System

- Tackle the Speech Synthesis/Voice Conversion attacks commonly referred to as **Logical Access** attacks.
  - Binary Classification task: **Feature Extractor** followed by a **Classifier** to give a result if a test speech utterance is bonafide or spoofed.
- Explore various feature extraction techniques such as MFCC, CQCC, Mel Spectrum and couple it with both DNN and Non Neural Network architectures to understand the performance of the resulting models.

Related Work

Model 1 [1]

STC antispoofing systems

- countermeasure system: LFCC(Linear Frequency Cepstral Coefficients)feature extractions from the first 600 frames
- Architecture : LCNN which uses normal kaiming initialization and softmax loss function

Model 1 Fusion method

Fusion of five subsystems:

- Subsystem 1,2 :LFCC-GMM
- Subsystem 3 : LCNN - CMVN (celepstral mean and variance normalisation)
- Subsystem 4 : log power spectrogram derived from CQT
- Subsystem 5 : log power spectogram derived from DFT

Related Work

Model 2 [2]

- countermeasure system: linear filterbank coefficients
- Architecture : ResNet-18

Model 2 Fusion method

Fusion of two subsystems:

- Subsystem 1: Resnet + linear filterbank coefficients
- Subsystem 2: Resnet + CQCC based features

Technical Approach -Setup

Dataset

Publicly available spoofing datasets

- ASVspoof 2017 v2.0
- ASVspoof 2019 LA [3]

The dataset comprises of:

- **Training set:** 25381, **Development set:** 24987, **Testing set:** 71934

Evaluation Metric

- Equal Error Rate (EER)
  - Decision threshold where the false acceptance and the false rejection rates are equal.
- Tandem Detection Cost Function (t-DCF) [4]
  - Takes into account both the ASV system error and CM system error into consideration.

Technical Approach - Extractor

Algorithm

Feature Extraction

- MFCC (Mel Frequency Cepstral Coefficients) - Available @ Librosa python
  - Windowing the signal
  - Log (DFT)
  - Warping the frequencies on a Mel Scale
  - Inverse DCT
- CQCC (Constant Q Cepstral Coefficients)
  - Constant Q transform (CQT), an alternative to Fourier based approaches to time-frequency analysis
  - In built python implementation not available, we will implement this in python based using the Matlab implementation as our reference.

Technical Approach - Classifier

Algorithm

Non NN models

- GMM (Gaussian Mixture Models) - Type of clustering method, each cluster is modified according to a different Gaussian distribution
- SVM (Support Vector Machines)
- VAEs (Variational Auto Encoders)

DNN models

- CNN - RNN - A combination of CNN along with sequence models such as RNN, GRU to capture temporal information present in the data.

Progress

- Implemented the CQCC pipeline in python.
  - Validated the output with the Matlab implementation and experiencing some bugs.
- Used MFCC feature extraction technique and using SVM as a classifier (in-progress)

Next Steps

- Explore the one class learning model suggested in [5] and employ it with DNN models.
- Extend the CM system to accommodate the Physical Access attacks (replay attacks) as well.

References

[1] Lavrentyeva, S. Novoselov, A. Tseren, M. Volkova, A. Gorlanov, and A. Kozlov, "STC antispoofing systems for the ASVspoof2019 challenge," in Proc. Interspeech, 2019, pp. 1033–1037.
[2] Chen, A. Kumar, P. Nagarsheth, G. Sivaraman, and E. Khoury, "Generalization of Audio Deepfake Detection," in Proc. Odyssey, 2020, pp. 132–137.
[3] Yamagishi, Junichi; Todisco, Massimiliano; Sahidullah, Md; Delgado, Héctor; Wang, Xin; Evans, Nicolas; Kinnunen, Tomi; Lee, Kong Aik; Vestman, Ville; Nautsch, Andreas. (2019). ASVspoof 2019: The 3rd Automatic Speaker Verification Spoofing and Countermeasures Challenge database, [sound]. University of Edinburgh. The Centre for Speech Technology Research (CSTR). https://doi.org/10.7488/ds/2555

References

[4] Kanervisto, Anssi Hautamäki, Ville Kinnunen, Tomi Yamagishi, Junichi. (2022). Optimizing Tandem Speaker Verification and Anti-Spoofing Systems.

[5] Y. Zhang, F. Jiang and Z. Duan, "One-Class Learning Towards Synthetic Voice Spoofing Detection," in IEEE Signal Processing Letters, vol. 28, pp. 937-941, 2021, doi: 10.1109/LSP.2021.3076358.