ECE 209AS - Project Presentation

Automatic Speech Verification - Spoof Detection

Sidarth Srinivasan (005629203)
Nithin Varma (005851269)

June 10, 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 (TTS)
    - Voice Conversion (VC)
- 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: Developed a CM system that consists of **Feature Extractor** followed by a **Classifier** to give a result if a test speech utterance is bonafide or spoofed.
- Explored various feature extraction techniques such as MFCC, CQCC, Mel Spectrum and coupled it with SVM classifier and also GMM generative classifier to understand the performance of the resulting models.

Related Work

Fusion Models

- Model 1 - Combines Feature extraction with Deep Models
  - Spec + ResNet
  - MFCC + ResNet
  - LFCC + ResNet
- Model 2 - Uses sequential models to extract features coupled with traditional ML classifiers
  - LC-GRNN + SVM
  - LC-GRNN + PLDA
  - LC-GRNN + LDA

Problem Setup

Main Question

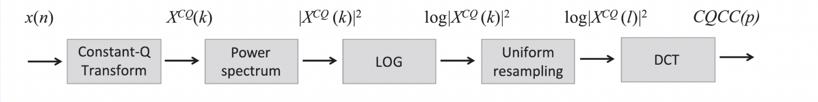
- How can the system defend against unknown spoofing attacks a.k.a **generalization ability** ?
  - **Intuition:** One class classification approach with modified loss function to shrink the embedding space of the target class.
- How to match single system performances to fusion model which are computationally expensive ?
  - **Intuition:** Using Generative models or Auto encoder models as an alternate to Deep fused models

Technical Approach - Extractor

Algorithm

Feature Extraction

- MFCC (Mel Frequency Cepstral Coefficients) - Available @ Librosa python.
- CQCC (Constant Q Cepstral Coefficients) - Implemented in python based on the block diagram below:



Technical Approach - Classifier

Models

GMM

- 3 GMMs of 144, 256 & 512 mixture components modules with expectation-maximization (EM) algorithm with random initialisation were trained.
  - Score for a given test occurrence is computed as the log-likelihood ratio as following :

$$\Lambda(X) = logL(X|\Theta_n) - logL(X|\Theta_s) \tag{1}$$

  - X - Test utterance feature vectors, L - Likelihood function, Θ_n - GMMs for bonafide speech, Θ_s - GMM for spoofed speech.

SVM

- 2 SVMs with mean-variance normalisation performed on the extracted features applied on a linear/RBF kernel and the default parameters of the Scikit-Learn library.

Technical Approach

Dataset & Protocols

- Publicly available ASVspoof 2019 LA [3] - Based on the VLTK corpus, a multi-speaker (46 male, 61 female) speech database.
  - **Training set:** 25380 with 2580 bonafide, 22800 spoofed utterances
  - **Development set:** 24987 with 2548 bonafide, 22296 spoofed utterances.
  - **Testing set:** 71934 with 7355 bonafide, 63882 spoofed utterances.
- Spoofed data is generated by using 17 TTS and VC algorithms.
  - 6 known spoofing systems with 2 VC and 4 TTS.
  - 11 unknown spoofing systems with unknown division.

Technical Approach (cont)

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.

Tandem detection cost function (t-DCF)

$$
\begin{aligned}
\text{t-DCF}(s,t) = {} & C_{\text{miss}}^{\text{asv}} \cdot \pi_{\text{tar}} \cdot P_{\text{a}}(s,t) \\
& + C_{\text{fa}}^{\text{asv}} \cdot \pi_{\text{non}} \cdot P_{\text{b}}(s,t) \\
& + C_{\text{fa}}^{\text{cm}} \cdot \pi_{\text{spoof}} \cdot P_{\text{c}}(s,t) \\
& + C_{\text{miss}}^{\text{cm}} \cdot \pi_{\text{tar}} \cdot P_{\text{d}}(s).
\end{aligned}
$$

  - C_{miss}^{asv} - Cost of ASV system rejecting a target trial.
  - C_{fa}^{asv} - Cost of ASV system accepting a non-target trial.
  - C_{miss}^{cm} - Cost of CM system rejecting a bonafide trial.
  - C_{fa}^{cm} - Cost of ASV system accepting a spoof trial.
  - π - Priori probabilities, P• - Error rates

Technical Approach

Platform

- Models were trained on Google Collab Pro on K80 and T4 GPUs with 32 GB RAM.
- Few of the pre-processing blocks were run on local machine.

Results - Development, Test

Model - Dev	**t-DCF**	**EER**
GMM - MFCC	0.0167	0.67
GMM - CQCC	0.0663	1.38
SVM - MFCC	0.0812	3.45
SVM - CQCC	0.0748	3.37

Model - Eval	**t-DCF**	**EER**
GMM - MFCC	0.2366	9.57
GMM - CQCC	0.2116	8.09
SVM - MFCC	0.3186	10.62
SVM - CQCC	0.3095	11.31

Retrospection

- **What worked?** - Adopting the one class learning approach helped in generalising the model for unknown spoof attacks.
- **What did not work?** - Although single systems did give comparable results to state of the art fusion models, better performance was expected. Probably a feature fusion could have aided in better results.
- **What could have been done differently?** - Using deep models to extract features rather than using MFCC, CQCC.
- **Future directions** - Exploring performances on individual spoof attacks and propose maybe an ensemble architecture to handle different spoofing attacks.

Work Split

- Construction of Data Pipeline - Nithin
- Data Preprocessing - Sidarth
- CQCC Implementation - Sidarth & Nithin
- GMM models - Sidarth & Nithin
- SVM models - Nithin
- Documentation - Sidarth

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.