Speaker Recognition

Software project progress report

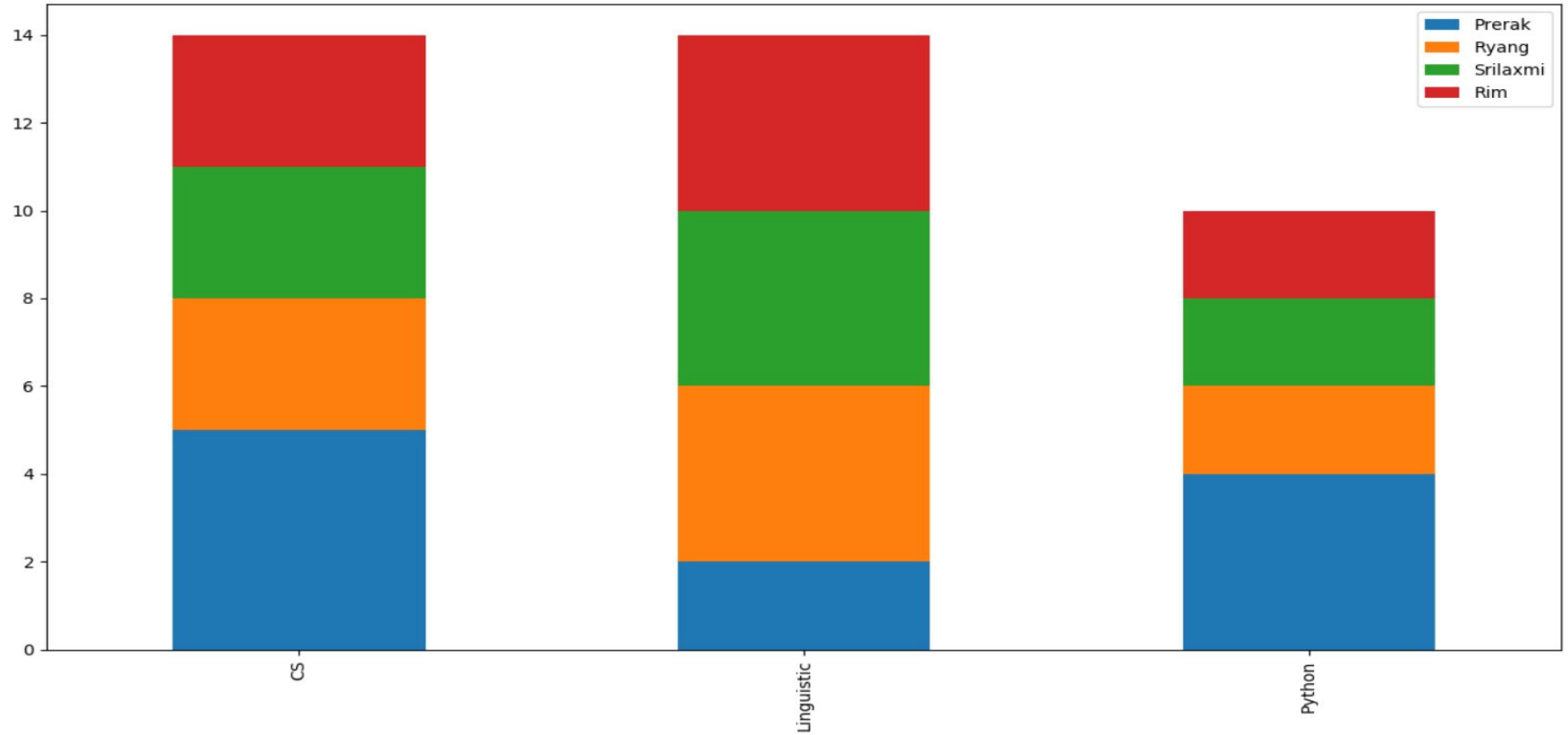
Abrougui Rim
Balard Srilakshmi
Srivastava Prerak
Yang Ruoxiao

October 2, 2019

UNIVERSITÉ DE LORRAINE

Institut des sciences du Digital
Management & Cognition

Background of Group Members



Outline

- Abstract
- Application
- Important Terms
- Baseline Models
- Objectives
- Our Approach
- Timelines
- Further Work

Abstract

- ***Popular technique***
- **Sub areas:**
  - Speaker identification
  - Speaker verification
- **Motivation**
- **Data:**
  - *Voxforge*
- **Purpose of our work:**
  - Train our model in French corpus for now
- **The method:**
  - Deep Neural Network Embeddings (i-vectors)
  - Transferring learning (kaldi tool)

Applications

- Secure access control by voice
  - Voice dialing, banking transaction over telephone, database access, multi-speaker tracking, etc.
- Verification of identity in e-commerce
- Make daily life more convenient

Important Terms

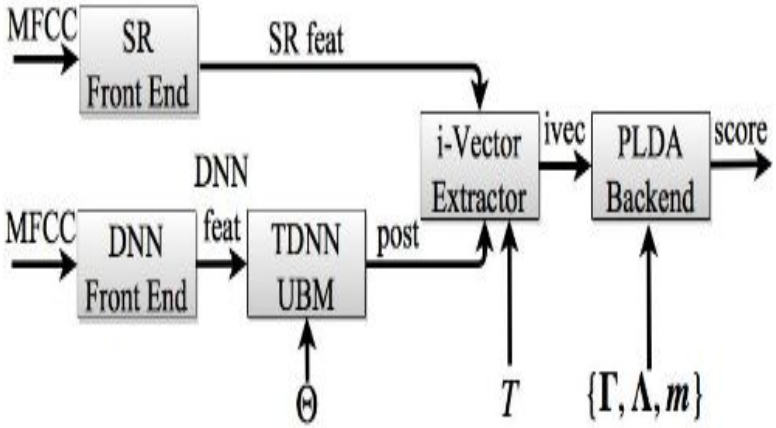
- Data Augmentation
- Variable-length Utterance
- Text Dependent and Text Independent
- In-Domain data and Out Of Domain Data
- Channel Variabilities
- Mel Frequency Cepstral Coefficient (MFCC)

Baseline (i-vector Model)

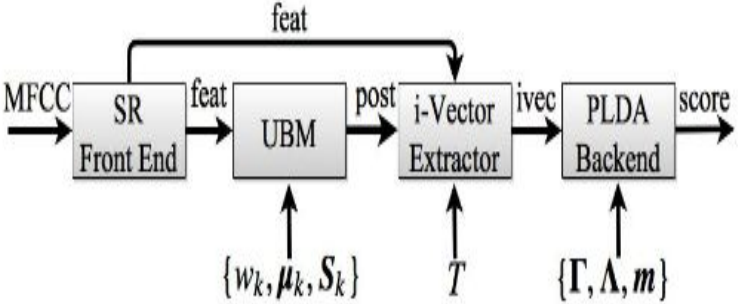
- It is based on Gaussian mixture model–universal background model (GMM-UBM)
- The I-vector is a compact representation that summarizes what is happening in a given speech recording
- it's a discrete spectrogram
- I-vector extraction is essentially a dimensionality reduction of the GMM supervector
- It factorizes the speech signal into the phonetic factor and the speaker factor

Comparison Between GMM-UBM v/s TDNN GMM

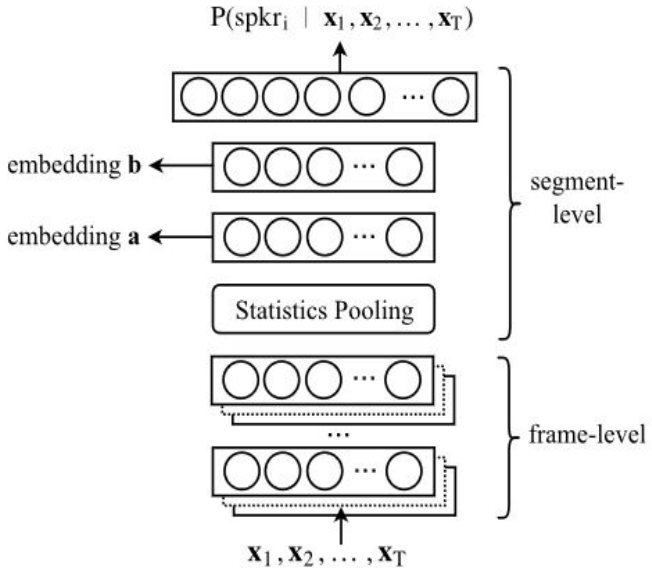
TDNN Based I-vector



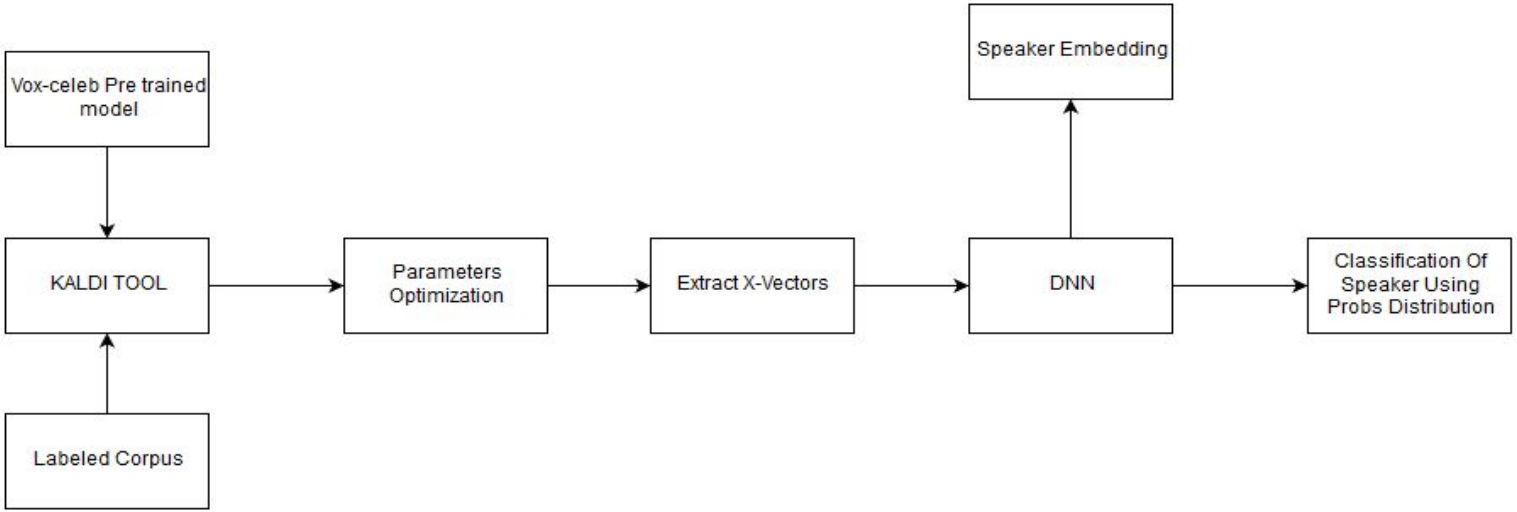
GMM-UBM



X-vectors


P(spkr$_i$ | $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$)
embedding **b**
embedding **a**
Statistics Pooling
segment-level
frame-level
$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$

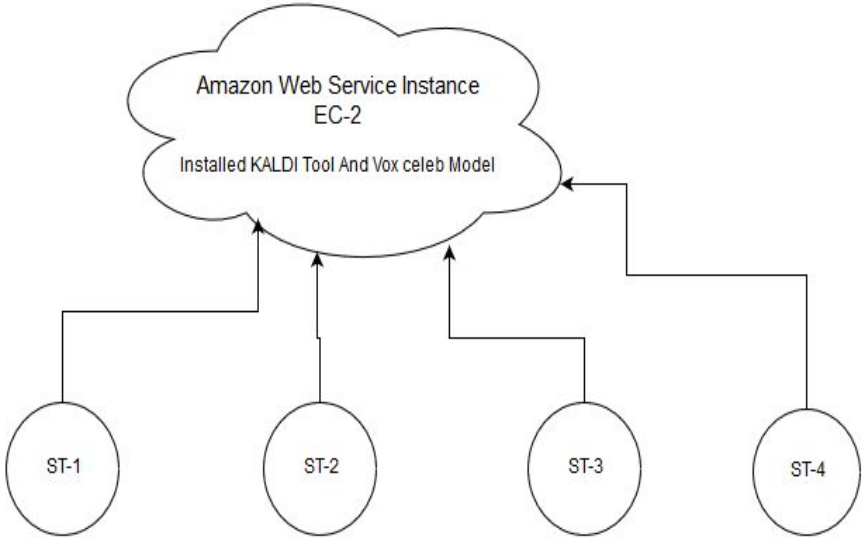
Our Approach



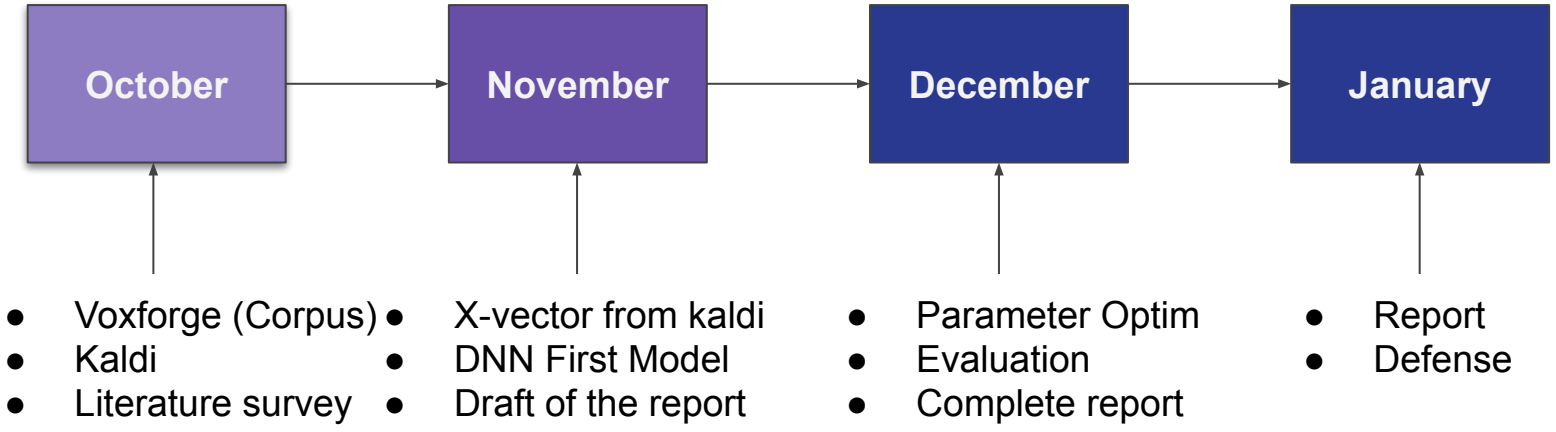
Evaluation

- **Precision:** how relevant are the positive detection
- **Recall:** probability of detection
- **Average accuracy**
- **F-score:** relation between data positive label and those given by classifier based on per-class average

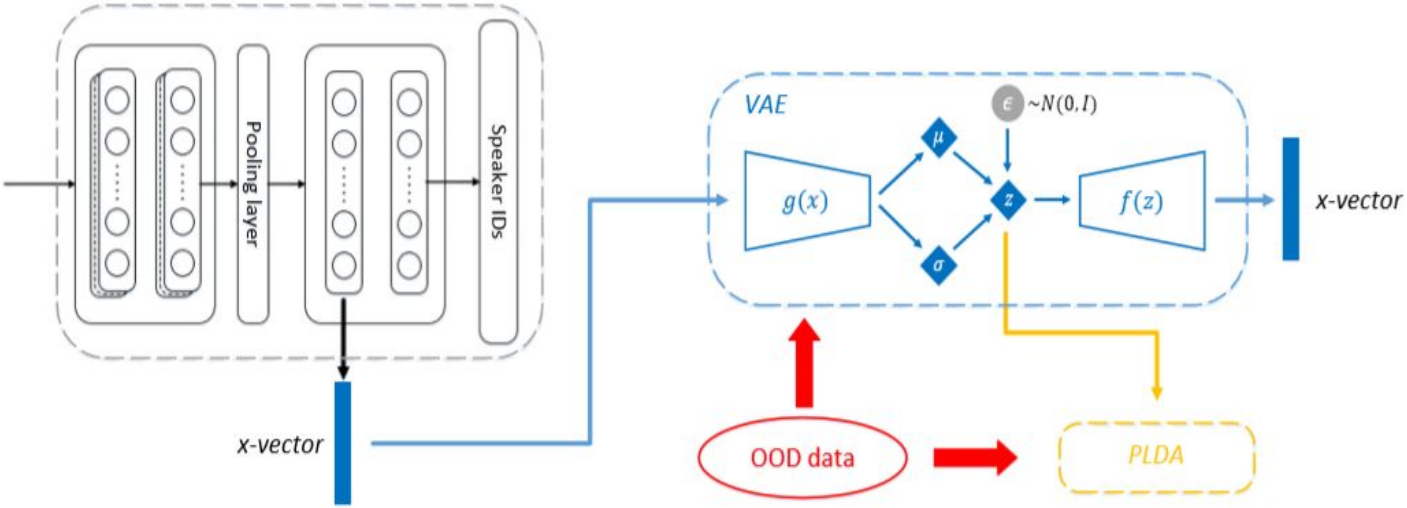
Kaldi Tool Robust Usage



Milestones


October
November
December
January
Voxforge (Corpus)
Kaldi
Literature survey
X-vector from kaldi
DNN First Model
Draft of the report
Parameter Optim
Evaluation
Complete report
Report
Defense

Future Work With Variational Autoencoder



GitLab Repo

https://gitlab.com/prerakshrivastava/asr-sv

The project structure of this repo as of now

- Data
- Models
- Utils
- Docs/Reports
- Kaldi
- Plots

Bibliography

- Snyder, David, Daniel Garcia-Romero, Gregory Sell, Daniel Povey and Sanjeev Khudanpur. “X-Vectors: Robust DNN Embeddings for Speaker Recognition.” *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018): 5329-5333.
- Snyder, David, Daniel Garcia-Romero, Daniel Povey and Sanjeev Khudanpur. “Deep Neural Network Embeddings for Text-Independent Speaker Verification.” INTERSPEECH (2017).
- Wang, X., Li, L., & Wang, D. (2019). VAE-based Domain Adaptation for Speaker Verification. ArXiv:1908.10092 [Cs, Eess]. Retrieved from http://arxiv.org/abs/1908.10092
- ……

Thank you!

Merci de votre attention!

Milestones

- **October:**
  - **Bibliography**
  - **Kaldi**
  - **Corpora**
- **November:**
  - **Neural network**
- **December:**
  - **Evaluation, realization and preparing the report and defense**
- **January:**
  - **Report submission and defense**