PROBING THE INFORMATION ENCODED IN X-VECTORS

Desh Raj, David Snyder, Daniel Povey, Sanjeev Khudanpur

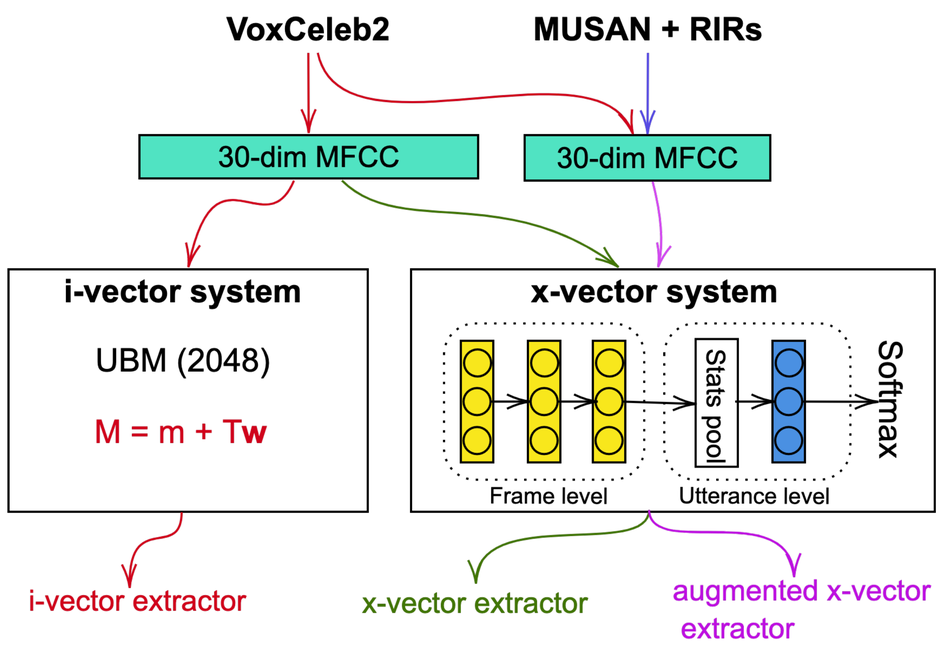
Center for Language and Speech Processing & Human Language Technology Center of Excellence
Johns Hopkins University, Baltimore, MD USA

Overview

- We probe speaker embeddings for information related to the channel, linguistic content, and meta information (utterance length, augmentation type).
- We analyze why augmentation helps for training x-vector extractors.

Speaker Embeddings

We train **i-vectors** and **x-vectors** (augmented and unaugmented) of dimensions 128, 256, 512, and 768.

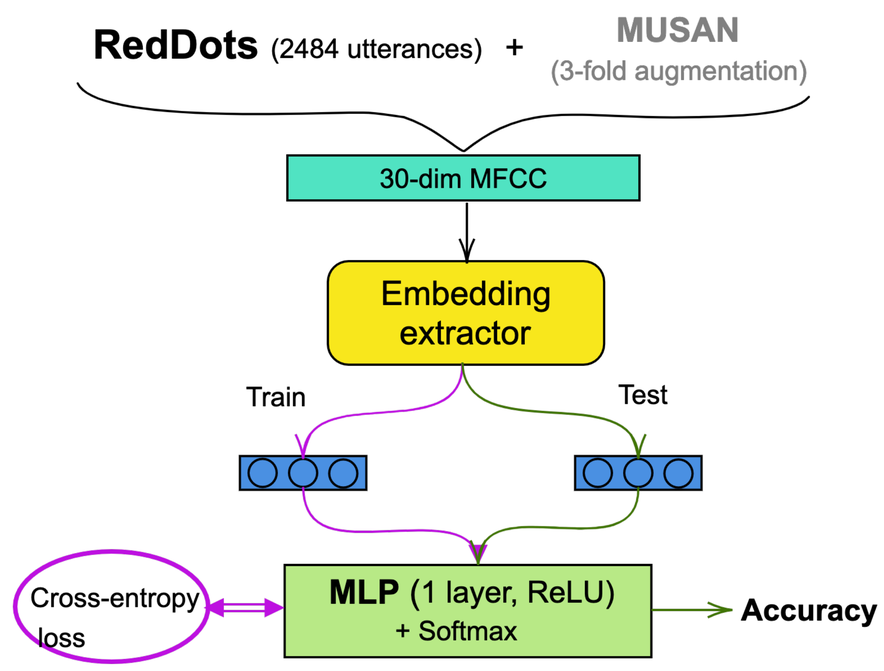


Probing Method

- **Better classification performance indicates that the embedding contains more information.**
- Classifier: MLP with single hidden layer (500-dimensional) and ReLU activation.

Speaker-related	Speaker gender, Speaking rate
Channel-related	Session id
Linguistic content	Transcription, Word recognition, Phoneme recognition
Meta-information	Augmentation type, Utterance length

Table: Probing tasks

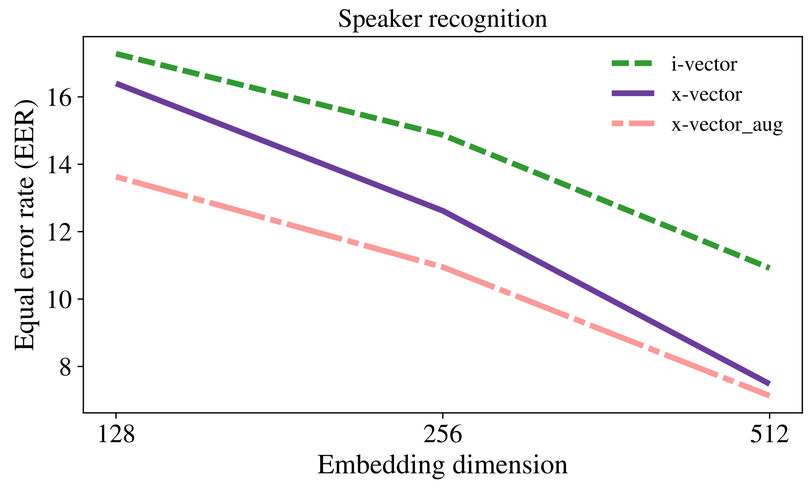


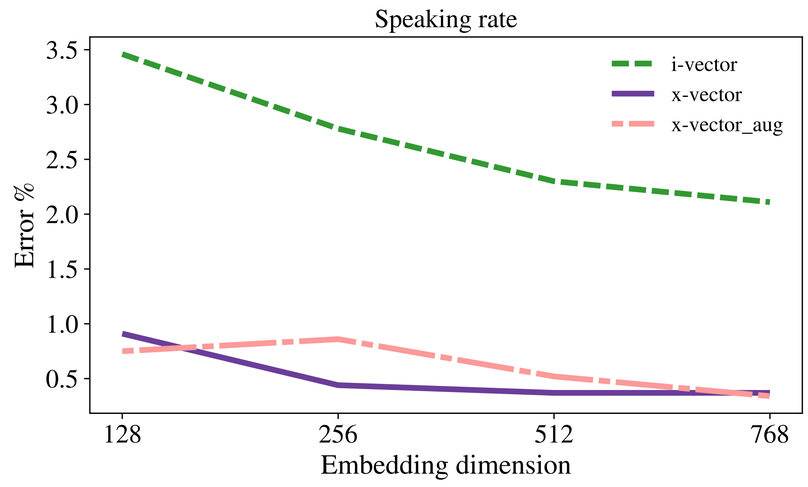
Contact Information

- Desh Raj
- Email: draj@cs.jhu.edu
- Website: https://desh2608.github.io

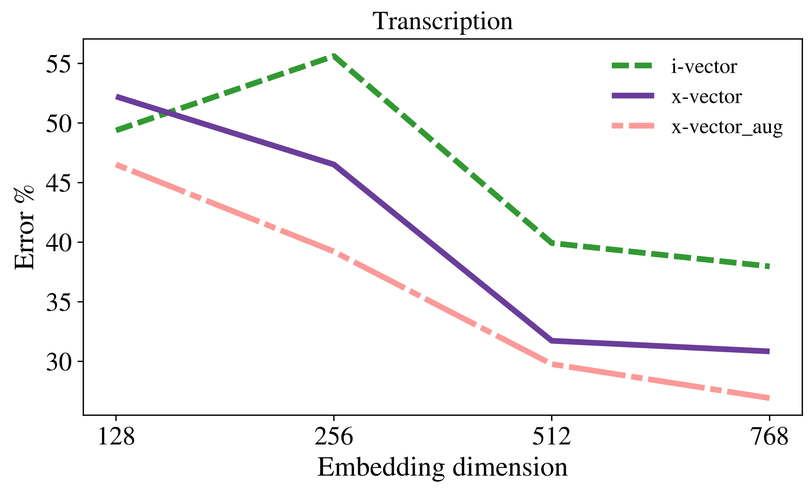
Results

- **Augmented x-vectors outperform i-vectors and unaugmented x-vectors on speaker recognition and related tasks (well known).**

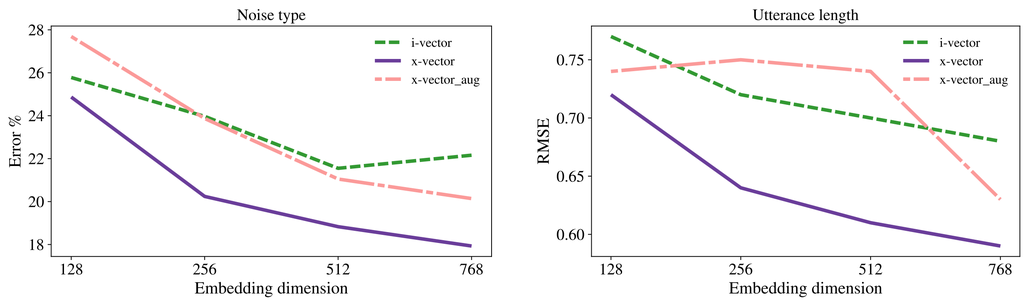




- **All speaker embeddings capture some information related to the lexical content in the utterance.**



- **Augmentation during extractor training trades some information such as noise type and utterance length for better speaker recognition performance.**



Conclusions

- Speaker embeddings contain information related to the **channel, linguistic content, and meta information**.
- Augmentation during extractor training trades some of this information (**noise type, utterance length**) for more speaker-related information.

Acknowledgements

This work was partially supported by NSF CRI Grant No 1513128, and DARPA LORELEI Contract No HR0011-15-2-0024.