Semi-supervised Nuisance-attribute Networks for Domain Adaptation

Weiwei LIN, Man-Wai MAK, Youzhi Tu and Jen-Tzung Chien

Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University

Dept. of Electrical and Computer Engineering, National Chiao Tung Univerity

Summary

- We propose semi-supervised a nuisance attribute networks (SNAN) to reduce the domain mismatch in i-vectors and x-vectors.
- The SNAN is based on the idea of nuisance attribute removal in inter-dataset variability compensation (IDVC).
- The architecture of SNAN allows us to incorporate the out-of-domain speaker labels into the semi-supervised training process through softmax loss, center loss and triplet loss.
- Using the SNAN as a preprocessing step for PLDA, we achieve a relative improvement of 11.8% in EER on NIST 2016 SRE compared to PLDA without adaptation.

X/I-Vector/PLDA Training

- We used the pre-trained DNN from the Kaldi repository.
- The i-vector system is based on gender-independent UBM 2048 mixtures and 600 dimensional total variability matrix. They were trained using SRE04–10 and Switchboard data with augmentation.
- X/I-vectors extracted from NIST 2004–2010 SREs were used to train gender-independent PLDA models.
- X/I-vector pre-processing:
  - Length-norm + LDA
  - 600 dim $\rightarrow$ 200 dim

Work flow of our approach

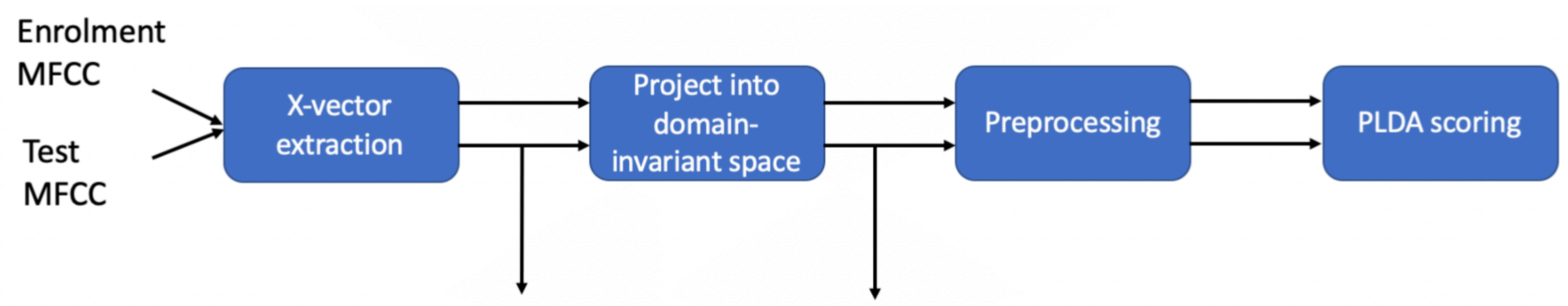


Fig. 1: T-SNE embedding of x-vectors from different datasets.

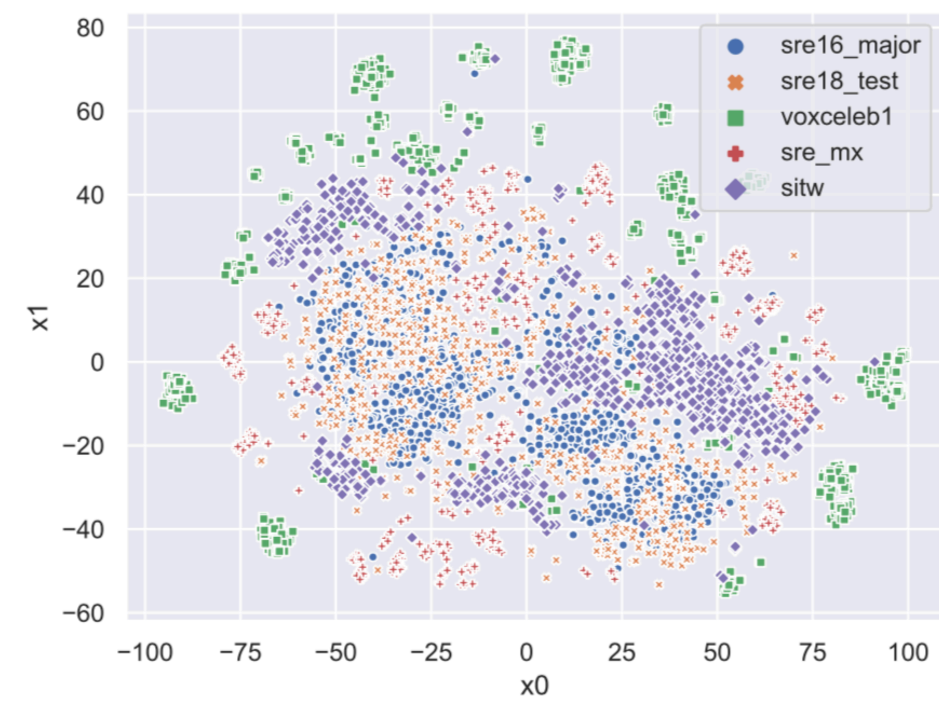
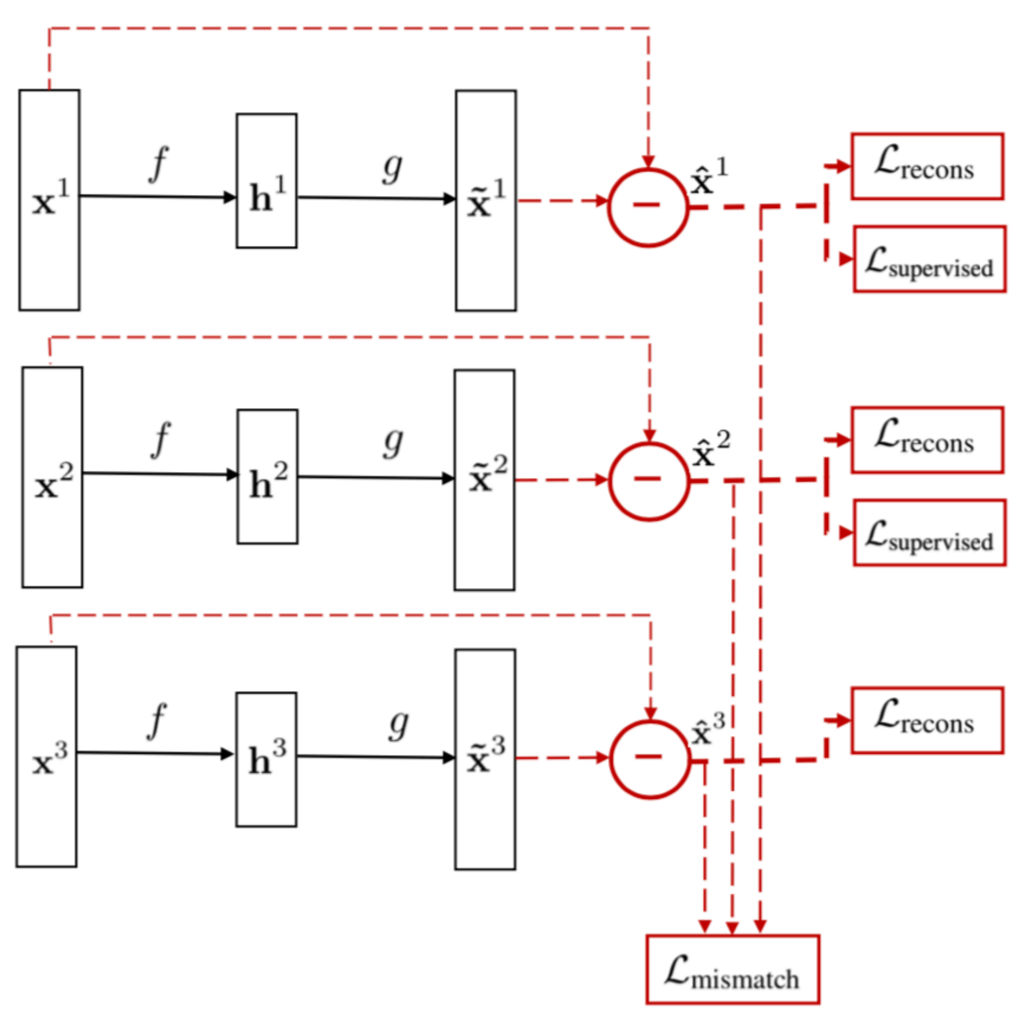


Fig. 2: T-SNE embedding of DAE transformed x-vectors from different datasets.

Semi-supervised Nuisance Attribute Network (SNAN)



- The network contains three objective functions.
- $\mathcal{L}_{\text{recons}}$ is a mean square error loss.
- $\mathcal{L}_{\text{supervised}}$ is a supervised loss. In this paper we tried softmax loss, center loss and triplet loss:

$$\mathcal{L}_{\text{center}} = \frac{1}{2}\sum_{i=1}^{B}\|\hat{\mathbf{x}}_i - \mathbf{c}_{y_i}\|^2$$

$$\mathcal{L}_{\text{triplet}} = \max\left\{\|\hat{\mathbf{x}}_a - \hat{\mathbf{x}}_p\|^2 - \|\hat{\mathbf{x}}_a - \hat{\mathbf{x}}_n\|^2 + m,\ 0\right\}$$

- $\mathcal{L}_{\text{mismatch}}$ is a domain-wise MMD defined as:

$$\mathcal{L}_{\text{mismatch}} = \sum_{d=1}^{D}\sum_{\substack{d'=1\\ d'\neq d}}^{D}\left(\frac{1}{N_d^2}\sum_{i=1}^{N_d}\sum_{i'=1}^{N_d}k(\hat{\mathbf{x}}_i^d, \hat{\mathbf{x}}_{i'}^d) - \frac{2}{N_d N_{d'}}\sum_{i=1}^{N_d}\sum_{j=1}^{N_{d'}}k(\hat{\mathbf{x}}_i^d, \hat{\mathbf{x}}_j^{d'}) + \frac{1}{N_{d'}^2}\sum_{j=1}^{N_{d'}}\sum_{j'=1}^{N_{d'}}k(\hat{\mathbf{x}}_j^{d'}, \hat{\mathbf{x}}_{j'}^{d'})\right)$$

- Total loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{mismatch}} + \alpha\mathcal{L}_{\text{recons}} + \beta\mathcal{L}_{\text{supervised}}$$

Performance on SRE16

I-vector systems

Methods	EER	mCprim	aCprim
No Adapt	12.78	0.74	0.94
IDVC	12.17	0.73	0.90
SNAN	11.95	0.72	0.87
SNAN (softmax)	11.61	0.71	0.87
SNAN (center loss)	11.76	0.72	0.86
SNAN (Triplet loss)	11.67	0.72	0.85

X-vector systems

Methods	EER	mCprim	aCprim
No Adapt	10.74	0.65	0.86
IDVC	11.24	0.65	0.89
SNAN	10.35	0.61	0.81
SNAN (softmax)	10.28	0.61	0.81
SNAN (center loss)	10.31	0.61	0.81
SNAN (Triplet loss)	10.57	0.62	0.8

Reference

1. Lin, Weiwei, et al. "Reducing domain mismatch by maximum mean discrepancy based autoencoders." *Proc. Odyssey*. 2018.
2. Lin, Wei-wei, Man-Wai Mak, and Jen-Tzung Chien. "Multisource i-vectors domain adaptation using maximum mean discrepancy based autoencoders." *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* 26.12 (2018): 2412-2422.
3. Snyder, David, et al. "X-vectors: Robust DNN embeddings for speaker recognition." *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.