InfoVDANN Supplementary Materials

Youzhi Tu, Man-Wai Mak and Jen-Tzung Chien

This document: http://to be added.pdf

May 14, 2020

Abstract

This document provides the derivation of the Monte Carlo estimate of mutual information, the statistical significance tests of the SRE performance, and the metric for selecting the dimension of LDA in the manuscript "Variational Domain Adversarial Learning with Mutual Information Maximization for Speaker Verification" submitted to IEEE/ACM Transactions on Audio, Speech and Language Processing.

1 Monte Carlo Estimate of Mutual Information

The mutual information (MI) between the latent variable $\mathbf{z}$ and the input $\mathbf{x}$ is defined as the KL divergence between the joint probability distribution $q_\phi(\mathbf{x}, \mathbf{z})$ and the product of their marginal distributions $p_{\mathcal{D}}(\mathbf{x})q_\phi(\mathbf{z})$

$$\begin{aligned} I_q(\mathbf{x};\mathbf{z}) &= \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{x},\mathbf{z})}{p_{\mathcal{D}}(\mathbf{x})q_\phi(\mathbf{z})}\right] \\ &= \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{q_\phi(\mathbf{z})}\right] \\ &= \mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log q_\phi(\mathbf{z}|\mathbf{x})] - \mathbb{E}_{q_\phi(\mathbf{z})}\left[\log q_\phi(\mathbf{z})\right] \\ &= -\mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\left[\mathbb{H}\left[q_\phi(\mathbf{z}|\mathbf{x})\right]\right] - \mathbb{E}_{q_\phi(\mathbf{z})}\left[\log q_\phi(\mathbf{z})\right]. \end{aligned} \tag{1}$$

In (1), $q_\phi(\mathbf{z}|\mathbf{x})$ is the variational posterior in the terminologies of variational autoencoders (VAEs), where ϕ parameterizes the encoder of a VAE; $q_\phi(\mathbf{z})$ is the aggregated posterior, i.e., $q_\phi(\mathbf{z}) = \int_{\mathbf{x}} p_{\mathcal{D}}(\mathbf{x})q_\phi(\mathbf{z}|\mathbf{x})d\mathbf{x}$.

The first term on the right-hand side of (1) is the average negative entropy of $\mathbf{z}$'s drawn from $q_\phi(\mathbf{z}|\mathbf{x})$. For a given $\mathbf{x}_s$, we assume $q_\phi(\mathbf{z}|\mathbf{x}_s) = \mathcal{N}\left(\mathbf{z}; \boldsymbol{\mu}_s, \text{diag}\left(\boldsymbol{\sigma}_s^2\right)\right)$, where $\boldsymbol{\mu}_s \equiv \boldsymbol{\mu}(\boldsymbol{x}_s;\phi)$ and $\boldsymbol{\sigma}_s \equiv \boldsymbol{\sigma}(\boldsymbol{x}_s;\phi)$ are the mean and standard deviation outputs of the encoder before the sampling process. Then the negative entropy can be analytically computed as follows:

$$-\mathbb{H}\left[q_\phi(\mathbf{z}|\mathbf{x}_s)\right] = -\frac{1}{2}\sum_{j=1}^{J}\left[1 + \log(2\pi) + \log\sigma_{sj}^2\right], \tag{2}$$

where J is the dimension of $\mathbf{z}$.

The second term $\mathbb{E}_{q_\phi(\mathbf{z})}[\log q_\phi(\mathbf{z})]$ can be estimated by Monte Carlo methods as in (16) in the manuscript:

$$\mathbb{E}_{q_\phi(\mathbf{z})}[\log q_\phi(\mathbf{z})] \approx \frac{1}{B}\sum_{s=1}^{B}\left[\log \frac{1}{B}\sum_{b=1}^{B} q_\phi\left(\mathbf{z}_s|\mathbf{x}_b\right)\right], \tag{3}$$

where B is the mini-batch size and $\mathbf{z}_s$ is a sample from the output of the InfoVAE's encoder given a sample $\mathbf{x}_b$ uniformly sampled from $p_{\mathcal{D}}(\mathbf{x})$, i.e., $\mathbf{z}_s$ is drawn from $\mathcal{N}\left(\mathbf{z}; \boldsymbol{\mu}_b, \mathrm{diag}\left(\boldsymbol{\sigma}_b^2\right)\right)$ (see Fig. 1 in the manuscript).

Finally, the MI is calculated as

$$\hat{I}_q(\mathbf{x};\mathbf{z}) = \frac{1}{B}\sum_{s=1}^{B}\left\{-\frac{1}{2}\sum_{j=1}^{J}\left[1+\log(2\pi)+\log\sigma_{sj}^2\right] - \log\frac{1}{B}\sum_{b=1}^{B} q_\phi\left(\mathbf{z}_s|\mathbf{x}_b\right)\right\}. \tag{4}$$

During the testing stage, B can be set to a large value (e.g., 1,024) for an accurate estimate of $I_q(\mathbf{x};\mathbf{z})$.

Alternatively, the MI can be expressed as

$$\begin{aligned} I_q(\mathbf{x};\mathbf{z}) &= \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log\frac{q_\phi(\mathbf{x}|\mathbf{z})}{p_{\mathcal{D}}(\mathbf{x})}\right] \\ &= \mathbb{E}_{q_\phi(\mathbf{z})}\mathbb{E}_{q_\phi(\mathbf{x}|\mathbf{z})}\left[\log q_\phi(\mathbf{x}|\mathbf{z})\right] - \mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\left[\log p_{\mathcal{D}}(\mathbf{x})\right] \\ &= -\mathbb{E}_{q_\phi(\mathbf{z})}\left[\mathbb{H}\left[q_\phi(\mathbf{x}|\mathbf{z})\right]\right] + \log N, \end{aligned} \tag{5}$$

where N is the number of test samples uniformly drawn from $p_{\mathcal{D}}(\mathbf{x})$. The term $\log N$ is resulted from the uniform sampling on the data set [1]. As entropy is always positive for random variables, the MI is bounded by $\log N$ from (5), which can provide a reference for the MI estimates.

2 Statistical Significance Tests

We conducted McNemar's tests [2] on the SRE performance to test the significance of the performance gain achieved by the InfoVDANN. As shown in Table 1, the P-values of the McNemar's tests between both InfoVDANNs and the others are mostly zeros for SRE16 and SRE18-CMN2. This means that the improvement of both InfoVDANNs over VDANN, DANN and the baseline is statistically significant. The "adp" in Table 1 represents the Kaldi's PLDA adaptation.

3 Selection of the LDA dimension

The dimension of LDA projection was determined by the EERs evaluated on the development sets. For SRE18-CMN2, we used the SRE18-CMN2 development

Table 1: P-values of the McNemar's tests [2]

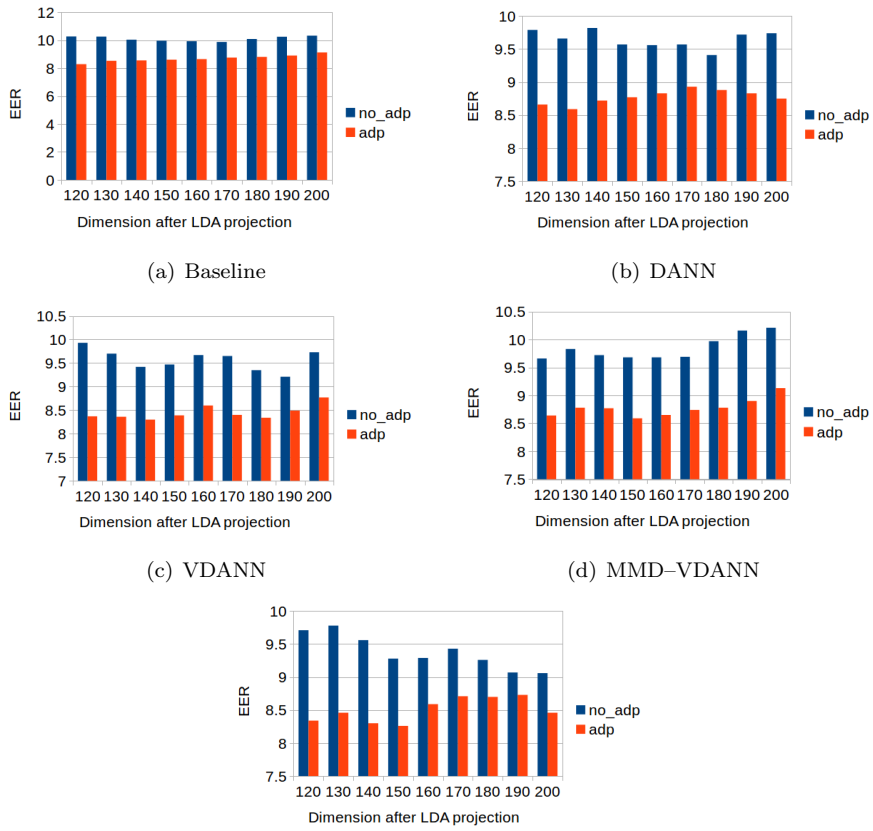
System1	System2	SRE16, All		SRE18-CMN2	
		w/o adp	w/ adp	w/o adp	w/ adp
MMD–VDANN	baseline	0	0	0	0
AAE–VDANN	baseline	0	0	0	0
MMD–VDANN	DANN	0	0	0	4.44×10^{-16}
AAE–VDANN	DANN	0	0	0	0
MMD–VDANN	VDANN	0	0	0	0
AAE–VDANN	VDANN	0	0	0	2.77×10^{-7}

set to compute the EERs. As shown in Fig. 1, for each system, the optimal dimension is 150 considering the EERs with and without PLDA adaptation. We thus set the dimension of the LDA projection to 150. We have a similar conclusion for SRE16, according to Fig. 2.

References

[1] M. Hoffman and M. Johnson, "ELBO surgery: yet another way to carve up the variational evidence lower bound," in *NIPS Workshop on Advances in Approximate Bayesian Inference*, 2016.

[2] L. Gillick and S. Cox, "Some statistical issues in the comparison of speech recognition algorithms," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, 1989, pp. 532–535.



(a) Baseline

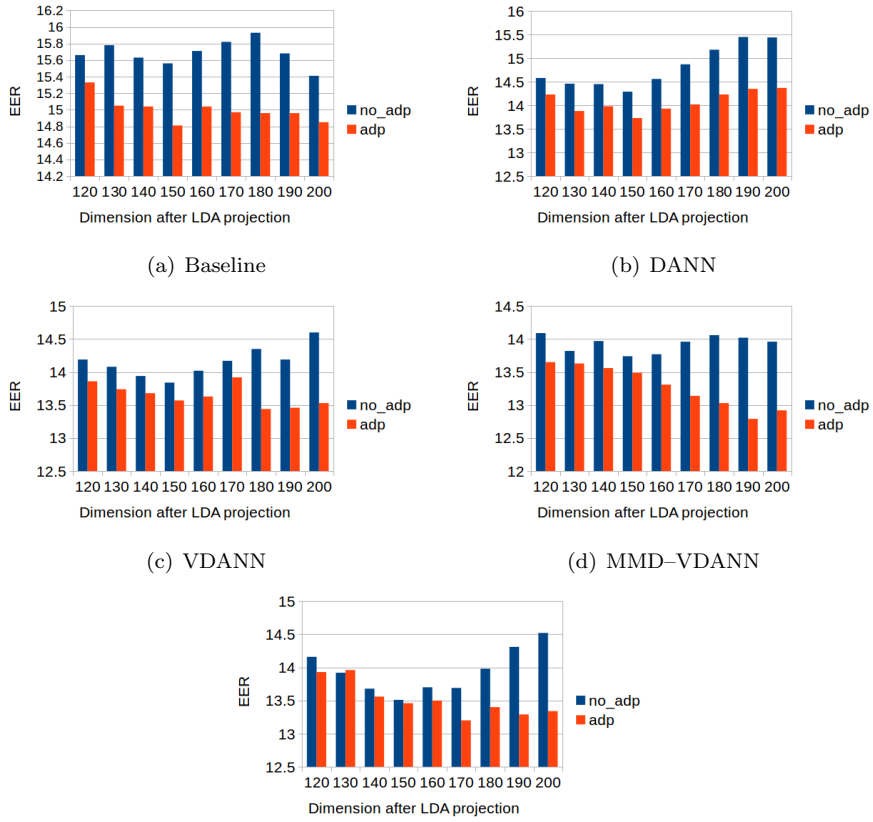
(b) DANN

(c) VDANN

(d) MMD–VDANN

(e) AAE–VDANN

Figure 1: EERs evaluated on the SRE18-CMN2 development set for different systems



(a) Baseline

(b) DANN

(c) VDANN

(d) MMD–VDANN

(e) AAE–VDANN

Figure 2: EERs evaluated on the SRE16 development set for different systems