**Target**: to investigate VAE-based unsupervised adaptation training(UAT) methods from different perspectives.

## Approach

|————front-end : Vox

|————back-end

|—— Vox

|—— VIVO

|—— UAT

Metrics:

PCA+Cosine PLDA

L-PLDA P-PLDA

V-L-PLDA V-P-PLDA

C-L-PLDA C-P-PLDA

## Data

|——Training data

|—— Vox

|—— VIVO

|—— Vox adapt VIVO

|——Test data

|—— VIVO

## Experiment

**setting**

c-vector: 200d

v-vcetor: 200d

LDA: 150d

PCA: 150d

**Baseline**

Method: Vox—— backend all trained on Vox

VIVO —— backend all trained on VIVO

Result:

| | Metric | PCA | V-PCA | C-PCA | PLDA | L-PLDA | P-PLDA | V-PLDA | V-L-PLDA | V-P-PLDA | C-PLDA | C-L-PLDA | C-P-PLDA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VOX | EER(%) | 18.37 | 21.00 | 16.03 | 18.51 | 14.82 | 14.58 | 16.72 | 15.97 | 16.66 | 15.58 | 14.29 | 14.66 |
| VIVO | EER(%) | 16.93 | 13.85 | 13.12 | 15.25 | 14.84 | 13.31 | 12.79 | 12.59 | 11.98 | 12.73 | 12.68 | 12.01 |

Conclusion:

（1）VIVO performs better than VOX because VIVO is in-domain test while Vox is out-of -domain test.

（2）VAE plays an important role in normalization when working with PLDA so the result gets better.

（3）when it comes to LDA, if it trained on Vox, it performs better than PCA. Once it trained on VIVO，vice versa. Since LDA is supervised model, it needs sufficient labels. But now we only have 40 people, so the result is not ideal. Meanwhile PCA is unsupervised model, under this circumstance it performs better.

**UAT**

1、adaptation of PLDA

Method: PLDA trained on Vox, and adapted on VIVO

LDA、VAE、CAE、PCA trained on Vox

Result:

| | Metric | PLDA | L-PLDA | P-PLDA | V-PLDA | V-L-PLDA | V-P-PLDA | C-PLDA | C-L-PLDA | C-P-PLDA |
|---|---|---|---|---|---|---|---|---|---|---|
| **VOX** | EER(%) | 18.51 | 14.82 | 14.58 | 16.72 | 15.97 | 16.66 | 15.58 | 14.29 | 14.66 |
| **VIVO** | EER(%) | 15.25 | 14.84 | 13.31 | 12.79 | 12.59 | 11.98 | 12.73 | 12.68 | 12.01 |
| **UAT** | EER(%) | 14.49 | 13.40 | 12.82 | 15.02 | 14.15 | 14.27 | 13.88 | 12.90 | 13.34 |

Conclusion:

（1）UAT does work since results of all the metrics are better than the one trained on Vox.

（2）Only PLDA L-PLDA P-PLDA outperform those trained on VIVO. It **may be** that VAE 、CAE is overfitting when training on Vox.

（3）CAE outperforms VAE, besides VAE even makes the result worse. Possible reason may be that(1) VAE is overfitting when trained on Vox (2)discrimination is more important than normalization under this circumstance ??

2、adaptation of VAE / CAE

Method:

UAT-VAE: VAE trained on Vox, adapted on VIVO, PCA、LDA、PLDA trained on VIVO

Result:

| | Metric | PLDA | L-PLDA | P-PLDA | V-PLDA | V-L-PLDA | V-P-PLDA | C-PLDA | C-L-PLDA | C-P-PLDA |
|---|---|---|---|---|---|---|---|---|---|---|
| **VOX** | EER(%) | 18.51 | 14.82 | 14.58 | 16.72 | 15.97 | 16.66 | 15.58 | 14.29 | 14.66 |
| **VIVO** | EER(%) | 15.25 | 14.84 | 13.31 | 12.79 | 12.59 | 11.98 | 12.73 | 12.68 | 12.01 |
| **UAT-PLDA** | EER(%) | 14.49 | 13.40 | 12.82 | 15.02 | 14.15 | 14.27 | 13.88 | 12.90 | 13.34 |
| **UAT-VAE** | EER(%) | —— | —— | —— | 13.10 | 13.07 | 12.50 | 13.07 | 13.13 | 12.65 |

Conclusion:

（1）Compared to Vox, UAT-VAE does work since results of all the metrics are better than the ones trained on Vox. So UAT-VAE works.

（2）Compared to VIVO, it did not outperform those trained on VIVO with VAE.

（3）Compared to UAT-PLDA , it outperforms generally. It may be indicated that VAE has a big impact on the effect of PLDA. To be more specific, the ability to normalize vectors is quite important in UAT. I think it may also be the expalation of (2). VAE training on VIVO is more precise than adaptation one when testing on VIVO.

（4）VAE/CAE has the ability to normalize vectors no matter whether applying UAT or not.

（5）CAE can make the data closer to their center, and from the table above we can see that it improves the performace of VAE. So utilizing label information is helpful.