I-Vector DNN Scoring and Calibration for Noise Robust Speaker Verification

Zhili TAN and Man-Wai MAK

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

Introduction

- Observing that adverse acoustic conditions and duration variability in utterances could have detrimental effect on PLDA scores, a number of score calibration methods have been proposed to compensated for the effect by modeling it as a shift in the PLDA scores.
- We propose to estimate the score shifts or the ideal clean scores by multitask DNNs using noisy i-vector pairs and their corresponding PLDA scores as input.
- Results based on noise contaminated speech in NIST 2012 SRE suggest that the multi-task DNNs can effectively calibrate the scores produced by a PLDA model, leading to superior performance as compared to the conventional linear calibration method.

Motivation

- Quality measure function (QMF)-based calibrated score:

$$S'_1 = w_0 + w_1 S + Q(\mathrm{SNR}_{tst}, \mathrm{SNR}_{tgt}) = w_0 + w_1 S + w_2 \mathrm{SNR}_{tst} + w_3 \mathrm{SNR}_{tgt}$$

where S is the uncalibrated score and $Q(\cdot,\cdot)$ is the approximated score shift.

- Ideal Score Shift $\delta_{score} \equiv \mathrm{PLDA}(\boldsymbol{x}_s^{cln}, \boldsymbol{x}_t^{cln}) - \mathrm{PLDA}(\boldsymbol{x}_s^{cln}, \boldsymbol{x}_t^{nsy})$
- However, the relationship between score shifts and utterances' SNR are fairly complex and definitely non-linear (see Fig. 2).
- At low SNR, the score shifts will become more difficult to estimate, which is a drawback of the methods that entirely rely on SNR of utterances.

DNN-Based Score Calibration

- *DNN Score Compensation*: Estimating Score Shifts by DNNs:

$$\mathrm{DNN}_A(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S) \approx \delta_{score}, \qquad S'_2 = S + \mathrm{DNN}_A(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S)$$

- *DNN Score Transformation*: Recovering Clean PLDA Scores by DNNs:

$$S'_3 = \mathrm{DNN}_B(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S) \approx S_{cln}$$

Multi-task DNNs for Score Calibration

- Having a single source of errors makes the backpropagation (BP) of error gradients very inefficient.
- One possible solution is to introduce some auxiliary tasks for the network to learn:

$$\mathrm{DNN}_C(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S) \approx \Big[\underbrace{[y_1, y_2]}_{\text{Classification}}, \underbrace{[S_{cln}, \delta_{score}, \mathrm{SNR}_i, \mathrm{SNR}_j, \mathrm{Dur}_i, \mathrm{Dur}_j]}_{\text{Regression}}\Big]$$

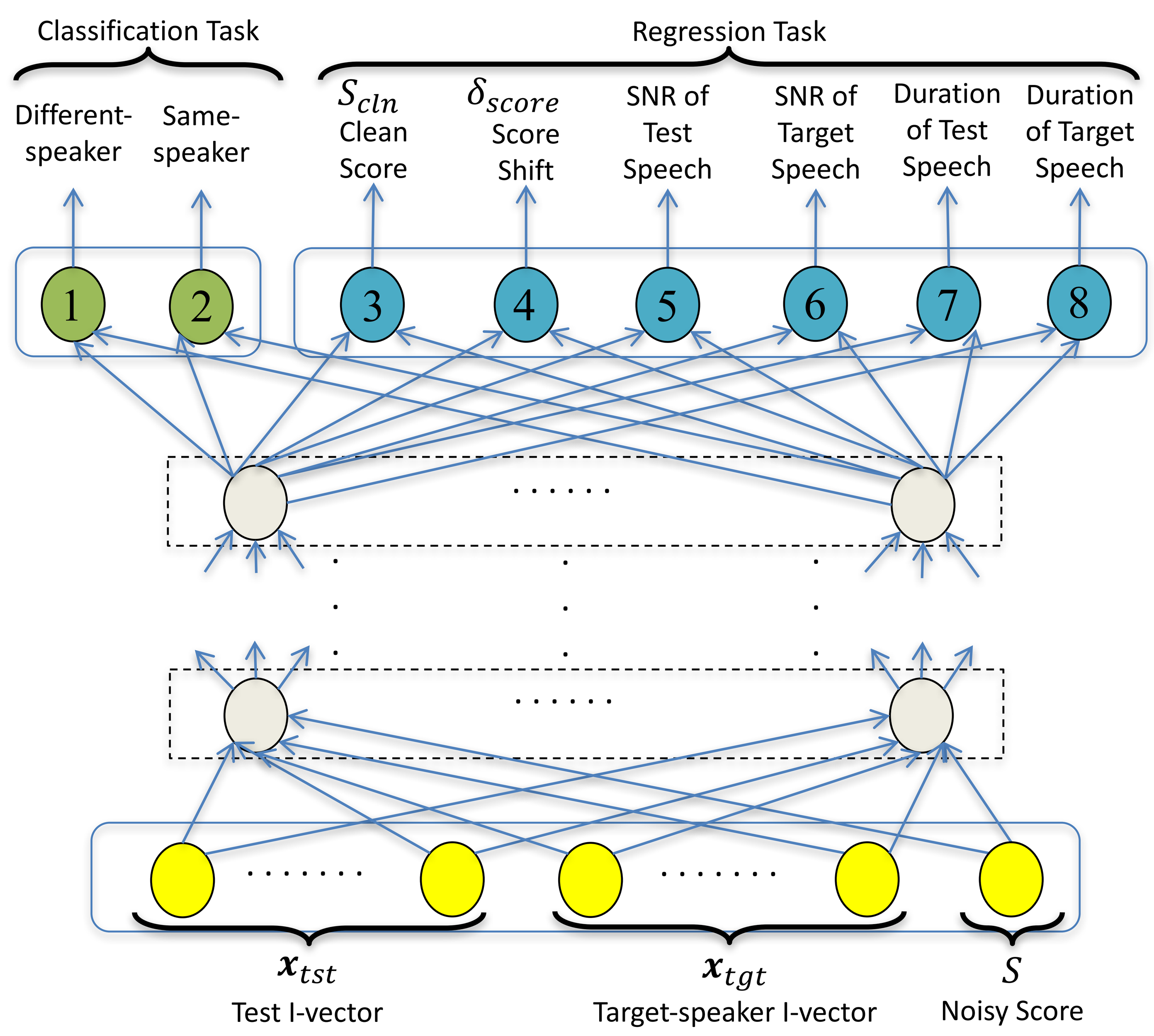

Fig. 1: Multitask DNN with classification and regression tasks.

- Recover Clean Scores by Multi-task DNN: $S'_4 = \mathrm{DNN}_C(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S)[3] \approx S_{cln}$
- Estimate Score Shifts by Multi-task DNN: $S'_5 = S + \mathrm{DNN}_C(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S)[4] \approx S_{cln}$
- Posterior odds by Multi-task DNN: $S'_6 = \log\left(\frac{\mathrm{DNN}_C(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S)[2]}{\mathrm{DNN}_C(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}, S)[1]}\right) = \log\left(\frac{p^+}{p^-}\right)$

DNN Scoring Machine

- A multi-task DNN without using the noisy PLDA scores as input.
- Advantage: the PLDA model is not necessary during the scoring stage, i.e., given an i-vector pair, we can obtain the approximated clean score or score shift from the DNN's outputs:

$$\mathrm{DNN}_D(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst}) \approx \Big[\underbrace{[y_1, y_2]}_{\text{Classification}}, \underbrace{[S_{cln}, \delta_{score}, \mathrm{SNR}_i, \mathrm{SNR}_j, \mathrm{Dur}_i, \mathrm{Dur}_j]}_{\text{Regression}}\Big]$$

- Recover Clean Scores by DNN scoring machine: $S'_7 = \mathrm{DNN}_D(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst})[3] \approx S_{cln}$
- Estimate Score Shifts by DNN scoring machine: $S'_8 = S + \mathrm{DNN}_D(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst})[4] \approx S_{cln}$
- Posterior odds DNN scoring machine: $S'_9 = \log\left(\frac{\mathrm{DNN}_D(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst})[2]}{\mathrm{DNN}_D(\boldsymbol{x}_{tgt}, \boldsymbol{x}_{tst})[1]}\right) = \log\left(\frac{p^+}{p^-}\right)$

Results on CC4 of NIST 2012 SRE (male)

- Test utterances are contaminated with different levels of babble noise.

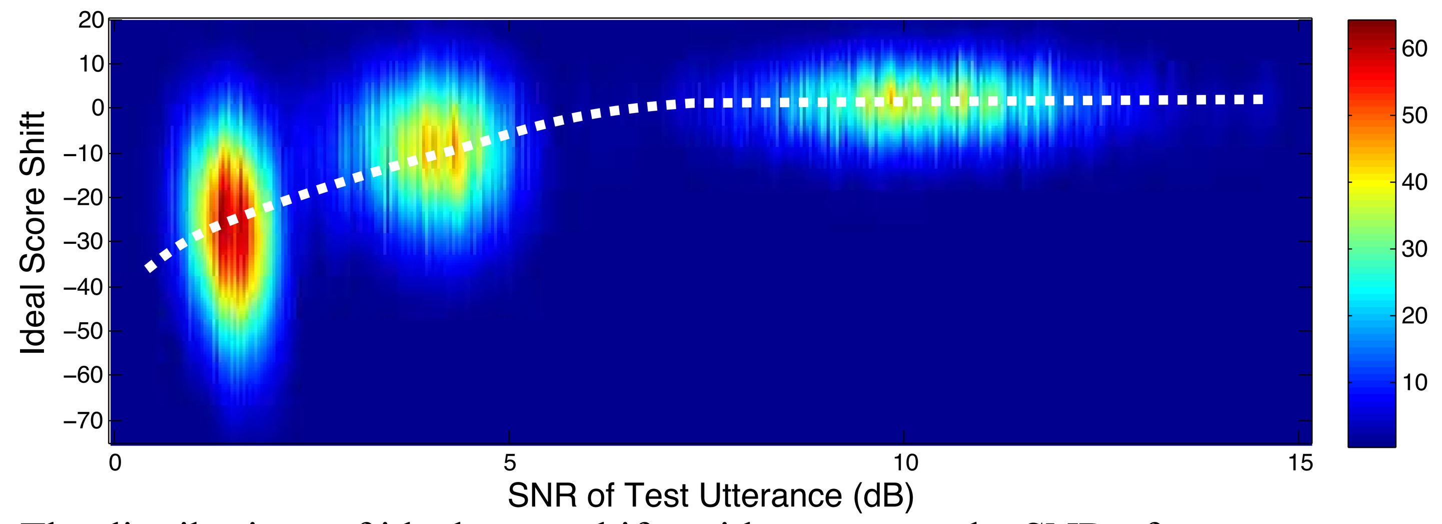

Fig. 2: The distributions of ideal score shifts with respect to the SNR of test utterances when the target-speaker utterances is clean.

- The uncalibrated PLDA scores play an important role in the calibration DNN.
- The classification task plays an important role in the training of the multi-task DNN.

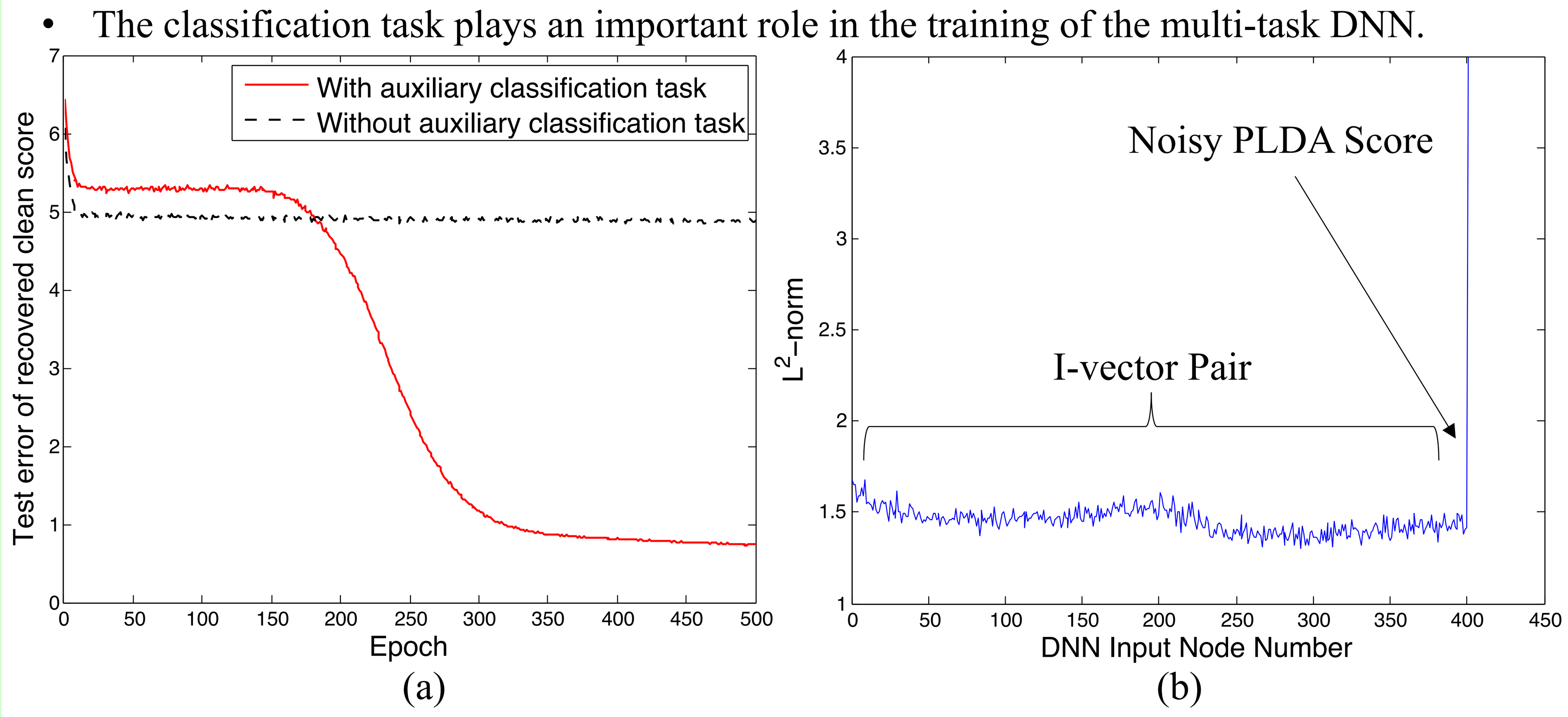

Fig. 3: (a) The mean squared test error between the recovered clean scores and the true clean score for 500 epochs of BP. (b) The l[2]-norm of the weight vectors in the bottom layer of Fig. 1. Each input node number corresponds to one weight vector representing the strength of that particular input to the first hidden layer.

- The DNN-based score calibrations outperform the conventional linear calibration method, and the SNR information improves the robustness significantly.
- All of the 3 calibration / scoring methods outperform the baseline:

Notation	Score Calibration Method	Original		6dB		0dB	
		EER(%)	minDCF	EER(%)	minDCF	EER(%)	minDCF
S'_1	Estimate SNR-dep Score Shift	1.68	0.209	2.28	0.269	5.35	0.754
S'_4	Recover Clean Scores by DNN	1.56	0.193	**2.21**	0.239	3.58	0.430
S'_5	Estimate Score Shifts by DNN	**1.54**	**0.192**	**2.21**	**0.238**	3.57	0.428
S'_6	Use Posterior Odds as Scores	1.70	0.193	2.23	0.245	**3.56**	**0.426**

Table 1: Performance of various DNN-based score calibration methods.

- Even without the noisy PLDA scores as input, the DNN is still able to estimate the score shift accurately:

Notation	Scoring Method	Score Calibration Method	Original		6dB		0dB	
			EER(%)	minDCF	EER(%)	minDCF	EER(%)	minDCF
S'_7	Recover Clean Score	N/A	2.51	0.308	3.02	0.349	**3.61**	**0.456**
S'_8	PLDA	Score Shift	**1.39**	**0.166**	**1.96**	**0.230**	3.80	0.571
S'_9	Posterior Odds	N/A	3.37	0.415	4.52	0.445	5.54	0.660

Table 2: Performance of multi-task DNNs without using the noisy PLDA scores as input.