Extracting Speaker-Specific Information with a Regularized Siamese Deep Network

Ke Chen and Ahmad Salman
School of Computer Science, The University of Manchester
Manchester M13 9PL, United Kingdom
{chen,salmana}@cs.manchester.ac.uk

Appendix

In this appendix, we derive the gradient of $L_D(X_1, X_2; \Theta)$ with respect to $\boldsymbol{u}_K(\boldsymbol{x}_{it})$ to obtain Eq. (5) in the main text.

To simplify the presentation, we first elucidate our notation system that is completely consistent to that used in the main text. We collectively denote the output of neurons in $\mathcal{CS}$ at the code layer or layer K of subnet i (*i=1,2*) as $\mathcal{CS}(X_i) = \left\{\left(\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j\right)_{j=1}^{|\mathcal{CS}|}\right\}_{t=1}^{T_B}$ for a speech segment of T_B frames, $X_i = \{\boldsymbol{x}_{it}\}_{t=1}^{T_B}$. Accordingly, we have $\boldsymbol{\mu}^{(i)} = \left(\mu_j^{(i)}\right)_{j=1}^{|\mathcal{CS}|}$ and $\Sigma^{(i)} = \left[\sigma_{ln}^{(i)}\right]$ $(l, n = 1, \cdots, |\mathcal{CS}|)$ where $\sigma_{ln}^{(i)} = \frac{1}{T_B - 1}\sum_{t=1}^{T_B}\left[\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_l - \mu_l^{(i)}\right]\left[\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_n - \mu_n^{(i)}\right]^T$. In the following derivation, we also drop all explicit parameters in $L_D(X_1, X_2; \Theta)$ and rewrite it into $L_D = L_m + L_S$ where $L_m = \mathcal{I}D_m + (1 - \mathcal{I})e^{-\frac{D_m}{\lambda_m}}$ and $L_S = \mathcal{I}D_S + (1 - \mathcal{I})e^{-\frac{D_S}{\lambda_S}}$.

Using our notation described above, we immediately achieve

$$\begin{aligned}
\frac{\partial L_D}{\partial \boldsymbol{u}_K(\boldsymbol{x}_{it})} &= \frac{\partial L_m}{\partial \boldsymbol{u}_K(\boldsymbol{x}_{it})} + \frac{\partial L_S}{\partial \boldsymbol{u}_K(\boldsymbol{x}_{it})} \\
&= \left(\left([\mathcal{I} - \lambda_m^{-1}(1 - \mathcal{I})e^{-\frac{D_m}{\lambda_m}}]\frac{\partial D_m}{\partial u_{Kj}(\boldsymbol{x}_{it})}\right)_{j=1}^{|\mathcal{CS}|}, (0)_{j=|\mathcal{CS}|+1}^{|\boldsymbol{h}_K|}\right) + \\
&\quad \left(\left([\mathcal{I} - \lambda_S^{-1}(1 - \mathcal{I})e^{-\frac{D_S}{\lambda_S}}]\frac{\partial D_S}{\partial u_{Kj}(\boldsymbol{x}_{it})}\right)_{j=1}^{|\mathcal{CS}|}, (0)_{j=|\mathcal{CS}|+1}^{|\boldsymbol{h}_K|}\right). \qquad \text{(A.1)}
\end{aligned}$$

To facilitate the presentation, we define $\psi_j(\boldsymbol{x}_{it}) = \frac{\partial D_m}{\partial u_{Kj}(\boldsymbol{x}_{it})}$ and $\xi_j(\boldsymbol{x}_{it}) = \frac{\partial D_S}{\partial u_{Kj}(\boldsymbol{x}_{it})}$. Now we simply need to calculate $\psi_j(\boldsymbol{x}_{it})$ and $\xi_j(\boldsymbol{x}_{it})$ for $j = 1, \cdots, |\mathcal{CS}|$ to obtain Eq. (5) in the main text.

As $D_m = ||\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}||_2^2 = \sum_{l=1}^{|\mathcal{CS}|}(\mu_l^{(1)} - \mu_l^{(2)})^2$, we have

$$\begin{aligned}
\psi_j(\boldsymbol{x}_{it}) &= \frac{\partial D_m}{\partial\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j}\frac{\partial\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j}{\partial u_{Kj}(\boldsymbol{x}_{it})} \\
&= \frac{\partial\sum_{l=1}^{|\mathcal{CS}|}(\mu_l^{(1)} - \mu_l^{(2)})^2}{\partial\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j}\frac{\partial\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j}{\partial u_{Kj}(\boldsymbol{x}_{it})} \\
&= p_j^{(i)}\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j\left[1 - \left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j\right], \qquad \text{(A.2)}
\end{aligned}$$

where

$$p_j^{(i)} = \frac{\partial\sum_{l=1}^{|\mathcal{CS}|}(\mu_l^{(1)} - \mu_l^{(2)})^2}{\partial\left(\mathcal{CS}(\boldsymbol{x}_{it})\right)_j} = \frac{2}{T_B}\text{sign}(1.5 - i)(\mu_j^{(1)} - \mu_j^{(2)}),$$

and

$$\frac{\partial\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j}{\partial u_{Kj}(\boldsymbol{x}_{it})} = \big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j\big[1-\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j\big]$$

given the fact that the transfer function used in the code layer or layer K is the sigmoid function. Collectively, we have $\boldsymbol{p}^{(i)} = \frac{2}{T_B}\text{sign}(1.5-i)(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})$.

Similarly, $D_S = ||\Sigma^{(1)}-\Sigma^{(2)}||_F^2 = \sum_{l=1}^{|\mathcal{CS}|}\sum_{n=1}^{|\mathcal{CS}|}(\sigma_{ln}^{(1)}-\sigma_{ln}^{(2)})^2$. Hence, we have

$$\begin{aligned}
\xi_j(\boldsymbol{x}_{it}) &= \frac{\partial D_S}{\partial\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j}\frac{\partial\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j}{\partial u_{Kj}(\boldsymbol{x}_{it})} \\
&= \frac{\partial\sum_{l=1}^{|\mathcal{CS}|}\sum_{n=1}^{|\mathcal{CS}|}(\sigma_{ln}^{(1)}-\sigma_{ln}^{(2)})^2}{\partial\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j}\frac{\partial\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j}{\partial u_{Kj}(\boldsymbol{x}_{it})} \\
&= q_j(\boldsymbol{x}_{it})\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j\big[1-\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_j\big], \qquad \text{(A.3)}
\end{aligned}$$

where

$$q_j(\boldsymbol{x}_{it}) = \frac{4}{T_B-1}\text{sign}(1.5-i)\sum_{n=1}^{|\mathcal{CS}|}(\sigma_{jn}^{(1)}-\sigma_{jn}^{(2)})\big[\big(\mathcal{CS}(\boldsymbol{x}_{it})\big)_n-\mu_n^{(i)}\big],$$

and, collectively, we have $\boldsymbol{q}(\boldsymbol{x}_{it}) = \frac{4}{T_B-1}\text{sign}(1.5-i)(\Sigma^{(1)}-\Sigma^{(2)})[\mathcal{CS}(\boldsymbol{x}_{it})-\boldsymbol{\mu}^{(i)}]$.

Inserting Eqs. (A.2) and (A.3) into Eq. (A.1), we obtain Eq. (5) in the main text as

$$\begin{aligned}
\frac{\partial L_D}{\partial \boldsymbol{u}_K(\boldsymbol{x}_{it})} &= \Big(\big([\mathcal{I}-\lambda_m^{-1}(1-\mathcal{I})e^{-\frac{D_m}{\lambda_m}}]\psi_j(\boldsymbol{x}_{it})\big)_{j=1}^{|\mathcal{CS}|},\big(0\big)_{j=|\mathcal{CS}|+1}^{|\boldsymbol{h}_K|}\Big) + \\
&\quad \Big(\big([\mathcal{I}-\lambda_S^{-1}(1-\mathcal{I})e^{-\frac{D_S}{\lambda_S}}]\xi_j(\boldsymbol{x}_{it})\big)_{j=1}^{|\mathcal{CS}|},\big(0\big)_{j=|\mathcal{CS}|+1}^{|\boldsymbol{h}_K|}\Big).
\end{aligned}$$