Multi-Task Autoencoder
For Noise-Robust Speech Recognition
Haoyi Zhang, Liu Conggui, Nakamasa Inoue, Koichi Shinoda
Tokyo Institute of Technology

Background

- Deep Denoising Autoencoder (DDAE) is often used as a preprocessor for speech recognition.
- We add another task, “deSpeech” from noisy speech, to form a multitask encoder.
- We expect it will enhance the denoising performance.

Multi-Task Autoencoder (MTAE)

- Objective function:
  \[ c \| s' - s \| + (1 - c) \| n' - n \| \]
  (We use \( c = 0.5 \))

There is no connection between green nodes and blue nodes.

Experiments

Input feature: MFCC 13 dim x 11 frames

Aurora 2 result (WER %)
Training: train, bubble, car, exhibition
Test Set A: train, bubble, car, exhibition
Test Set B: restaurant, street, airport, station
Test Set C: train, street with MIRS filter (others G.712)

<table>
<thead>
<tr>
<th>Multi-Task AutoEncoder (MTAE)</th>
<th>MFCC</th>
<th>ETSI -AFE</th>
<th>NMCC</th>
<th>DDAE</th>
<th>UNI</th>
<th>SHARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean</td>
<td>20dB</td>
<td>15dB</td>
<td>10dB</td>
<td>5dB</td>
<td>0dB</td>
<td>-5dB</td>
</tr>
<tr>
<td>Set A</td>
<td>0.7</td>
<td>1.0</td>
<td>1.3</td>
<td>2.3</td>
<td>5.0</td>
<td>17.9</td>
</tr>
<tr>
<td>Set B</td>
<td>0.7</td>
<td>1.3</td>
<td>2.5</td>
<td>7.3</td>
<td>20.1</td>
<td>48.0</td>
</tr>
<tr>
<td>Set C</td>
<td>0.7</td>
<td>1.1</td>
<td>2.2</td>
<td>5.4</td>
<td>10.9</td>
<td>36.8</td>
</tr>
<tr>
<td>Ave.</td>
<td>0.7</td>
<td>1.1</td>
<td>2.0</td>
<td>5.0</td>
<td>12.0</td>
<td>34.2</td>
</tr>
</tbody>
</table>

CHiME3 result (WER %)
Training: café, bus, street, pedestrian area
Test Set: real noisy data “et05_real”
Averaged over 0-20dB

<table>
<thead>
<tr>
<th>MFCC</th>
<th>DDAE</th>
<th>MTAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>et05_real</td>
<td>50.8</td>
<td>29.7</td>
</tr>
</tbody>
</table>

Summary

- Extracting noise help extracting speech in noisy speech.
- The partly-shared structure is effective in multi-task learning.
- The use of raw or log spectral features may be simpler and more effective.