Progress Report

Joe Frankel

Listen Meeting, 26th March 2007
Tandem ASR feature generation

- Context window
- Linear dimension-ality reduction
- Non-linear mapping
- Acoustic features
- Tandem feature

- Input window introduces contextual information
- Non-linear mapping trained to maximize separation between phone classes

Joe Frankel Progress Report
Tandem ASR feature generation

- Input window introduces contextual information
- Non-linear mapping trained to maximize separation between phone classes
- Linear dimensionality reduction
- Phone–discriminant non-linear mapping
Tandem ASR feature generation

- input window introduces contextual information
- non-linear mapping trained to maximize separation between phone classes
Construct MLP features for:

- Meeting domain farfield signals
- Mandarin broadcast news (BN)
- ...and see if MLPs previously trained on 2000hrs of continuous telephone speech (CTS) could be of benefit.
Construct MLP features for:

- meeting domain farfield signals
Construct MLP features for:

- meeting domain farfield signals
- Mandarin broadcast news (BN)
Task

Construct MLP features for:
- meeting domain farfield signals
- Mandarin broadcast news (BN)
- ...and see if MLPs previously trained on 2000hrs of continuous telephone speech (CTS) could be of benefit.
Method - Meetings

Three approaches taken for farfield meetings data:
Three approaches taken for farfield meetings data:

- train on farfield data from random initialization.
Three approaches taken for farfield meetings data:

- train on farfield data from random initialization.
- train on farfield data using CTS MLPs as initialization.
Three approaches taken for farfield meetings data:

- train on farfield data from random initialization.
- train on farfield data using CTS MLPs as initialization.
- train on farfield data from random initialization in a multi-task setting. Secondary task is speech enhancement mapping from farfield to nearfield data.
Rationale:
Rationale:

- using a shared representation, related tasks can act as a prompts for each other
Rationale:

- using a shared representation, related tasks can act as a prompts for each other
- reduced risk of over-training, as error function local minima for multiple tasks are unlikely to fall at the same location.
WER on NIST RT05s MDM data.

<table>
<thead>
<tr>
<th></th>
<th>MLP train data</th>
<th>word error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no MLP</td>
<td>N/A</td>
<td>36.7</td>
</tr>
<tr>
<td>random initialization</td>
<td>farfield</td>
<td>35.2</td>
</tr>
<tr>
<td></td>
<td>farfield MTL</td>
<td>34.3</td>
</tr>
<tr>
<td>adapted from CTS</td>
<td>nearfield</td>
<td>41.2</td>
</tr>
<tr>
<td></td>
<td>farfield</td>
<td>33.2</td>
</tr>
<tr>
<td></td>
<td>nearfield + farfield</td>
<td>33.0</td>
</tr>
</tbody>
</table>
MLPs were trained from a random initialization on the target data. Additionally, the following approaches to adaptation from English CTS MLPS were used:
MLPs were trained from a random initialization on the target data. Additionally, the following approaches to adaptation from English CTS MLPS were used:

- retain input-hidden layer, re-initialize hidden-output layer
MLPs were trained from a random initialization on the target data. Additionally, the following approaches to adaptation from English CTS MLPs were used:

- retain input-hidden layer, re-initialize hidden-output layer
- add additional (4th) layer to the CTS MLPs
# Mandarin MLP training CV accuracies

<table>
<thead>
<tr>
<th>initialization</th>
<th>Cross-validation accuracy (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tandem</td>
<td>HATS</td>
</tr>
<tr>
<td>random 5% free params</td>
<td>74.1</td>
<td>75.5</td>
</tr>
<tr>
<td>random 10% free params</td>
<td>74.8</td>
<td>76.2</td>
</tr>
<tr>
<td>adapted, 3-layer</td>
<td>76.8</td>
<td>77.2</td>
</tr>
<tr>
<td>adapted, 4-layer</td>
<td>75.5</td>
<td>76.5</td>
</tr>
</tbody>
</table>
Results - Mandarin broadcast news

WER on dev-04 and eval-04 data sets.

<table>
<thead>
<tr>
<th></th>
<th>word error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dev-04</td>
</tr>
<tr>
<td>no MLP</td>
<td>9.5</td>
</tr>
<tr>
<td>Chinese (70 hours)</td>
<td>8.2</td>
</tr>
<tr>
<td>Chinese (440 hours)</td>
<td>8.0</td>
</tr>
<tr>
<td>adapted (70 hours)</td>
<td>7.7</td>
</tr>
</tbody>
</table>
Tandem features well proven, though
Tandem features well proven, though
- mappings trained against phone targets
Tandem features well proven, though

- mappings trained against phone targets
- a data-driven set of target classes based on HMM states should be more useful in subsequent modelling.
Tandem features well proven, though

- mappings trained against phone targets
- a data-driven set of target classes based on HMM states should be more useful in subsequent modelling.
- confusable states which lead to common word confusions could be placed in different clusters.
Clustering

Current state of play:

- Performed state-level alignment of meetings data (AMI, ICSI, NIST, CMU, ISL)
- Mapped these down to the 3907 unique shared states found in the AMI ASR system
- About to generate state similarity matrix for clustering.
- HMM state GMMs and data available, so considering measures of the form for the similarity of model $k$ and $l$:
  \[
  \frac{1}{N_k} \sum_{y \in Y_k} p(y|\Theta_l) + \frac{1}{N_l} \sum_{y \in Y_l} p(y|\Theta_k)
  \]
Current state of play:

- performed state-level alignment of meetings data (AMI, ICSI, NIST, CMU, ISL)
Current state of play:

- performed state-level alignment of meetings data (AMI, ICSI, NIST, CMU, ISL)
- mapped these down to the 3907 unique shared states found in the AMI ASR system
Current state of play:

- performed state-level alignment of meetings data (AMI, ICSI, NIST, CMU, ISL)
- mapped these down to the 3907 unique shared states found in the AMI ASR system
- about to generate state similarity matrix for clustering.

Joe Frankel
Progress Report
Clustering

Current state of play:

- performed state-level alignment of meetings data (AMI, ICSI, NIST, CMU, ISL)
- mapped these down to the 3907 unique shared states found in the AMI ASR system
- about to generate state similarity matrix for clustering.
- HMM state GMMs and data available, so considering measures of the form for the similarity of model $k$ and $l$:

\[
\frac{1}{N_k} \sum_{y \in Y_k} p(y|\Theta_l) + \frac{1}{N_l} \sum_{y \in Y_l} p(y|\Theta_k)
\]