Phonetic decision tree in Janus

Based on the paper from Michael Finke and Ivica Rogina
“Wide Context Acoustic Modeling in Read vs. Spontaneous Speech”
Train a context dependent (CD) system with Janus

• Create context independent (CI) decision tree (simple script)
• Train CI system
• Add polyphone tree to CI - decision tree
• Train the distribution of polyphone tree
• Cluster polyphones → CD – decision tree
• Do training with the CD - decision tree
Trained CI system

Is center phone = iy
Add the polyphone tree (PTree)

Go through all the training data and collect all possible contexts that occur in the leaf node.

Possible source of problems: You have to specify what the maximum context size should be. It should match the questions you want to use.
Train the distribution of PTree
Cluster criterion

\[ p_i^l = \frac{1}{\gamma^l} \sum_{m \in L} \gamma_m \alpha_{m_i}, \quad \gamma^l = \sum_{m \in L} \gamma_m \]

\[ p_i^r = \frac{1}{\gamma^r} \sum_{m \in R} \gamma_m \alpha_{m_i}, \quad \gamma^r = \sum_{m \in R} \gamma_m \]

\[ D(q) = \gamma^l H^l + \gamma^r H^r - \gamma H \]

\[ -H^l = \sum p_i^l \log p_i^l \]

\[ -H^r = \sum p_i^r \log p_i^r \]

Question splits distribution into \( L \) and \( R \)

Counts of component \( i \) in model \( m \)

Counts of model \( m \)
Example: Question q splits context

q = “Is left context is t?”

q = Yes

q = No

\[ H = \sum p_i \log p_i \]

\[ H^t = \sum p^t_i \log p^t_i \]

\[ H^r = \sum p^r_i \log p^r_i \]
Example: Question q splits context

q = "Is left context t?"

Yes

No

Split only if enough data available for training the models

Stop after maximum number of splits is performed or no split is possible

Thomas Schaaf

interACT at CMU
Remove PTree and split codebooks

$q = \text{“Is left context is } t\text{?”}$

Yes

No

Create new distribution for each leave

Now train with the CD tree!
It is possible to do the same procedure with the CD system.
Why should we do so?
Results over context width

<table>
<thead>
<tr>
<th>Task</th>
<th>Context ±1</th>
<th>Context ±2</th>
<th>Context ±3</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSJ</td>
<td>20.9% WE</td>
<td>20.2% WE</td>
<td>19.9% WE</td>
</tr>
<tr>
<td>SWB</td>
<td>46.0% WE</td>
<td>43.6% WE</td>
<td></td>
</tr>
</tbody>
</table>
Main differences between HTK and Janus

• Distance:
  – HTK: Measures improvement in log-likelihood. Uses a single Gaussian to do so.
  – Janus: Maximum Entropy gain on the mixture weight distribution.

• Parameter reduction:
  – HTK: agglomerative clustering after splitting
  – Janus: do clustering in two steps
Some questions not discussed

• What is a useful context width?
• How many Gaussians should I use in the CI system?
• Can I do merge an split training for the CI system?
• What questions should I ask?
Two Stage Clustering

a) Codebook Clustering

b) Distribution Clustering