DECISION TREE DISTRIBUTION TYING BASED ON A DIMENSIONAL SPLIT TECHNIQUE

Authors: Heiga Zen
Keiichi Tokuda
Tadashi Kitamura

Presented by: Heiga Zen

Department of Computer Science,
Nagoya Institute of Technology, Japan

ICSLP 2002
Introduction

Phonetic Decision Tree based State Tying Approach

- Reduce free parameters in a system
- Generate unseen models
- All dimensions have the same sharing structure

All dimensions should have the same sharing structure?

Dimensional Split Phonetic Decision Tree

- Construct proper sharing structure of each dimension
- Capture correlations among dimensions
Split $S$ into $S_{q+}$ and $S_{q-}$ by question $q$, the difference of DL value, $\Delta_q$ is given by

$$
\Delta_q = \frac{1}{2} \left\{ \Gamma(S_{q+}) \log |\Sigma_{S_{q+}}| + \Gamma(S_{q-}) \log |\Sigma_{S_{q-}}| - \Gamma(S) \log |\Sigma_S| \right\} + K \log \Gamma(S_0)
$$

$K$ : Dimensionality of Feature Vector
$\Sigma$ : Covariance Matrix of Each Cluster
$\Gamma(\cdot)$ : State Occupancy Count for Each Cluster

$$
q_{\text{best}} = \left\{ q \left| \arg \min_{q \in Q} \Delta_q, \Delta_q < 0 \right. \right\}
$$

$q_{\text{best}} = \emptyset$ □ stop
Feature Dependent Phonetic Decision Tree

- Feature Dependent SSS [Matsuda, et al.; 2000]

  - Feature Dependent Phonetic Decision Tree

  ![Feature Dependent Phonetic Decision Tree Diagram]

  - 1st dimension
  - 2nd dimension
  - \( k-a+b \)
  - \( t-a+n \)
  - \( \ldots \)
  - \( K \)-th dimension

  - \( R_{\text{pre}} \)
  - \( R_{\text{wide}} \)
  - \( R_{\text{narrow}} \)
  - \( L_{\text{silence}} \)
  - \( R_{\text{narrow}} \)
  - \( L_{\text{pre}} \)
  - \( R_{\text{narrow}} \)
Problem in Feature Dependent Approach

State Tying Approach

Feature Dependent Approach
Dimensional Split Phonetic Decision Tree
**Dimensional Split Technique**

### $q_1$
- $H_{q_1} = \{1, 3, 4\}$
- $\Delta_{q_1} = \Delta_{q_1}^1 + \Delta_{q_1}^3 + \Delta_{q_1}^4$
- $\Delta_{q_1} = -0.8$

### $q_2$
- $H_{q_2} = \{\emptyset\}$
- $\Delta_{q_2} = 0$

### $q_3$
- $H_{q_3} = \{1, 6\}$
- $\Delta_{q_3} = \Delta_{q_3}^1 + \Delta_{q_3}^6$
- $\Delta_{q_3} = -1.5$

---

### Best Question

$q_{best} = \arg \min_{q \sqsubset Q} \Delta_q$

= $q_3$

### Dimensional Split

$S = \{1, 2, 3, 4, 5, 6\}$

$S_1 = \{1, 6\}$

$S_2 = \{2, 3, 4, 5\}$
Constructed Decision Tree by Dimensional Split Approach

Root Cluster

- Dimensional split
- All features

- **R_pre舌根**
  - YES
    - Dimensional split
    - Other features
    - R_narrow
      - YES
      - Dimensional split
      - Other features
      - R_narrow
      - NO
    - Dimensional split
    - Other features
    - R_halfnarrow
      - YES
      - Dimensional split
      - Other features
      - R_halfnarrow
      - NO
  - NO

- Dimensional split
- Other features
- R_narrow
- NO

- Dimensional split
- Other features
- R_halfnarrow
- NO
## Experimental Conditions

<table>
<thead>
<tr>
<th>Database</th>
<th>ATR continuous speech database B-set, 6 male speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data</td>
<td>5 male speakers, phonetically balanced 450 sentences</td>
</tr>
<tr>
<td>Test Data</td>
<td>53 sentences of remaining 1 male speaker</td>
</tr>
<tr>
<td>Test Method</td>
<td>Jack-knife approach</td>
</tr>
<tr>
<td>Recognition</td>
<td>Continuous phoneme recognition</td>
</tr>
</tbody>
</table>
# Clustering Results

## Number of leaf nodes and free parameters

<table>
<thead>
<tr>
<th>Method</th>
<th>#leaf nodes</th>
<th>#parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDT</td>
<td>3,044</td>
<td>344,624</td>
</tr>
<tr>
<td>FD-PDT</td>
<td>136,592</td>
<td>273,182</td>
</tr>
<tr>
<td>DS-PDT</td>
<td>93,838</td>
<td>311,916</td>
</tr>
</tbody>
</table>

**PDT** : Phonetic Decision Tree  
**FD-PDT** : Feature Dependent Phonetic Decision Tree  
**DS-PDT** : Dimensional Split Phonetic Decision Tree

All of these methods were based on MDL criterion.
Example of Sharing Structure in DS-PDT

1st state of phoneme "py"

The same color regions correspond to the same leaf node.

PDT clustered these 4 states into 2 states,
DS-PDT clustered these into 16 sub-distributions,
FD-PDT clustered these into 103 scalar distributions.
Example of Sharing Structure in DS-PDT

1st state of phoneme "py"

The same distribution is shared (17/56)
These features were independent from contextual environment in this task?
Number of Distributions in Each Dimension

Graph showing the number of distributions in each dimension for Mel-cepstrum, ΔMel-cepstrum, and ΔΔMel-cepstrum. The graph compares PDT, FD-PDT, and DS-PDT.
Experimental Results

Average Phoneme Error Rate (%) vs Number of Free Parameters

- PDT
- FD-PDT
- DS-PDT

1-mix 2-mix 4-mix 8-mix

Number of Free Parameters $10^5$
**Conclusion**

- DS-PDT can construct proper sharing structure for each dimension.

- DS-PDT achieved about 8% error reduction over conventional PDT.

**Future Work**

- Experiment on large corpus.

- Large vocabulary continuous word recognition.
Computational Cost

Machine: Pentium 4, 1.6 GHz
Decoder: HTK Viterbi Decoder
(modified for proposed acoustic models, not tuned)

- PDT: 7.0 x RT
- DS-PDT: 41.3 x RT
- FD-PDT: 59.5 x RT
Difference of DL in Each Dimension

The difference of Description Length in PDT based on MDL

\[ \Delta_q = \frac{1}{2} \left\{ \Gamma(S_{q+}) \log |\Sigma_{S_{q+}}| + \Gamma(S_{q-}) \log |\Sigma_{S_{q-}}| - \Gamma(S) \log |\Sigma_S| \right\} + K \log \Gamma(S_0) \]

In case of diagonal covariance,

\[ \Sigma_S = \text{diag} \left[ \sigma_{S,(1)}^2, \sigma_{S,(2)}^2, \ldots, \sigma_{S,(K)}^2 \right] \]

\[ \Delta_q = \sum_{k=1}^{K} \Delta_q^{(k)} \]

\[ \Delta_q^{(k)} = \frac{1}{2} \left\{ \Gamma(S_{q+}) \log \sigma_{S_{q+},(k)}^2 + \Gamma(S_{q-}) \log \sigma_{S_{q-},(k)}^2 - \Gamma(S) \log \sigma_{S,(k)}^2 \right\} + \log \Gamma(S_0) \]
Special Case of DS-PDT = State Tying Approach

Root Cluster

R_pre_tongue

YES NO

R_narrow

YES NO

R_half_narrow

φ

Dimensional split

all features
Special Case of DS-PDT = Feature Dependent Approach

Root Cluster

dimensional split
c(1)

R_posttongue

YES

NO

c(2)

other features

dimensional split

R_o

YES

NO

c(3)

other features

dimensional split

R_posttongue

YES

NO

c(4)

other features

R_o

YES

NO