Providing Sublexical Constraints for Word Spotting within the ANGIE Framework

Raymond Lau and Stephanie Seneff

{ raylau, seneff }@sls.lcs.mit.edu
http://www.sls.lcs.mit.edu

Spoken Language Systems Group
MIT Laboratory for Computer Science
Cambridge, Massachusetts
United States of America
Outline

• ANGIE
• Wordspotter
• Filler models
• Results
• Conclusions
What is ANGIE?

• Flexible, multipurpose system for speech processing
• Framework introduced in Seneff, Lau & Meng (ICSLP ’96)
• Word substructures characterized jointly by:
  – Context free grammar
  – Probabilistic model
• Possible applications include:
  – Flexible/extensible speech recognition tasks
  – Bidirectional letter/sound generation
  – Prosodic modeling
• Benefits include:
  – Pooling of data due to hierarchical structure
  – Generalization of knowledge to new words
  – Easy experimentation with subword representations
• Very regular layered structure
• Regular structure imposed by CF rules with lhs and rhs on adjacent layers
• Layers are sentence, word, morphology, syllabification, phonemics, phonetics
• No stress layer -- instead, distributed amongst layers
• Parsing proceeds left-to-right with each column built bottom-up
• Last two layers capture phonological variation
• Context dependencies typical in phonology learned by probability model
• Probabilities:
  • Terminal advancement
  • Bottom up trigram
Current Task: Wordspotting

• Task: Wordspot 39 city names in ATIS
  – Training 5000 utts, testing Dec ’93 test set
  – Similar task to Manos and Zue (ICASSP ’97)
• Objectives:
  – Explore effects of varying subword lexical model
    * Easy to do within the ANGIE framework
  – Further establish empirically the feasibility of using ANGIE for speech recognition tasks
  – Use as a natural foundation for building a full ANGIE speech recognizer
Wordspotter

- Start with segment based graph as in MIT's SUMMIT
- Use mixture diagonal Gaussian acoustic models for context-independent phones:
  - MFCC means averaged over thirds of segments
  - MFCC derivatives across segment boundaries
- Perform left-to-right search of phone graph
  - Partial ANGIE parses computed for partial theories
    * Well supported by ANGIE's left-to-right bottom-up parsing strategy
  - Best ANGIE parse score used as linguistic score
Search Strategy

• Previous work with ANGIE used best-first strategy
  – Proved inadequate empirically for wordspotting
  – Possible reason: difficulties in normalizing short vs. long theories for comparison
• Current strategy: Variant of stack decoder
  – c.f., Jelinek (IEEE ’76), Paul (ICASSP ’91)
  – Extend all paths at the earliest unexplored time boundary based on score
  – Prune based on a maximum number of paths permitted at any boundary
Filler Models

- ANGIE provides subword lexical model for the filler space
- Different ANGIE configurations give us a range of models
- Start with least constraint: phone bigram
- End with most constraint: full ANGIE layered model with 1200 word lexicon
- In all cases, no cross-word constraints (e.g., word $n$-gram) used
Range of Filler Models

- Phones
  - Only phone bigram used
- Pseudo-words (e.g., flid: f l ih dcl d)
  - “Invent” possible pseudo-words bottom-up
- Syllables (e.g., ciscofran: s ih s kcl k uh f axr n)
  - Syllable is highest unit
  - Syllable ordering not enforced
- Morphs (e.g., conflighting: kcl k aa n f l ay tcl t iy ng)
  - Syllables with ordering enforced
- Known words plus pseudo-words
  - 1200 words plus allow “invention” of pseudo-words
- Known words only
  - 1200 words
Results

<table>
<thead>
<tr>
<th>Filler Model</th>
<th>Figure of Merit</th>
<th>Rel. Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone Bigram</td>
<td>85.3</td>
<td>-</td>
</tr>
<tr>
<td>Pseudo-words</td>
<td>86.3</td>
<td>1.00</td>
</tr>
<tr>
<td>Syllables</td>
<td>87.7</td>
<td>0.56</td>
</tr>
<tr>
<td>Morphs</td>
<td>88.4</td>
<td>0.79</td>
</tr>
<tr>
<td>Words + Pseudo-words</td>
<td>88.6</td>
<td>0.79</td>
</tr>
<tr>
<td>Words</td>
<td>89.3</td>
<td>0.74</td>
</tr>
</tbody>
</table>

• More constraint leads to higher FOM
• Speed increases with constraints
  • Possible explanation: lower branchout
  • Exception: Syllables very fast
• Word bigram gets 93.9 FOM
Other Points

• Increasing subword lexical constraints on filler model improves performance
  – Another example of “full recognition is best”
  – Permitting pseudo-words in addition to known words did not help, even if vocabulary lowered to 400 words

• Integration of Chung’s ANGIE-based duration model improves performance even more (up to 91.6 FOM)
  – To be presented: W2C.3 (Wed, 12:30, Delphi)

• Test set coverage is 92% with vocab size of 400 and 86% with vocab size of 200 (where performance starts to swap relative ordering)
Future Work

- ANGIE is a workable framework for speech processing
  - Especially for research in subword lexical modeling
  - But also can leverage off of parse tree structure for acoustic modeling
- Natural next step is full speech recognition
  - Easy to do dynamic vocabulary updates
- Other tasks
  - Pronunciation server (integrates well with dynamic vocabulary recognizer)