Abstract

We report in detail the decoding strategy that we used for the past two DARPA Rich Transcription evaluations (RT’03 and RT’04) which is based on finite-state automata (FSA). We discuss the format of the static decoding graphs, the particulars of our Viterbi implementation, the lattice generation and the likelihood evaluation. Experimental results are given on the EARS database (English conversational telephone speech) with emphasis on our faster than real-time system.

Static decoding graphs

They are acceptors (instead of transducers) Arcs in English have three different types of labels:

- Words of option.
- Null states mark heavy incoming arcs labeled by
- e.g. due to LM back-off states.
- Emission states for which all incoming arcs are labeled by
- the same leaf label.
- Emission states for which all incoming arcs are labeled by
- words.
- Other.

Two different types of states:

- Emission labels (e.g. due to LM back-off states).
- Word labels.
- Other.

Every state keeps track of the N-best distinct word sequences arriving at

Lattice generation

Running beam pruning: pruning based on current max-

Successor look-up table: maps static to dynamic state

Graph memory layout: graph stored as a linear array of

Viterbi search speed-ups

Likelihood computation

Graph memory layout

Successor look-up table

Running beam pruning

Graph memory layout

Two different types of states:

- Emission labels (e.g. due to LM back-off states).
- Word labels.
- Other.

Every state keeps track of the N-best distinct word sequences arriving at

Static decoding graphs

<table>
<thead>
<tr>
<th>Number of leaves</th>
<th>Number of words</th>
<th>Number of n-grams</th>
<th>Number of states</th>
<th>Number of arcs</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.9K</td>
<td>32.9K</td>
<td>3.9M</td>
<td>18.5M</td>
<td>44.5M</td>
</tr>
<tr>
<td>21.5K</td>
<td>32.9K</td>
<td>4.2M</td>
<td>26.7M</td>
<td>68.7M</td>
</tr>
</tbody>
</table>

Search statistics:

Word error rate:

- RT03: 19.0%
- RT04: 16.4%

Lattice link density:

- RT03: 29.4
- RT04: 45.1

Lattice degree:

- RT03: 10
- RT04: 15

N-best degree:

- RT03: 5
- RT04: 5

Speaker-adapted decoding

- RT03: 16.1%
- RT04: 13.0%

LM rescoring + consensus

- RT03: 17.4%
- RT04: 14.5%

VM1.1 phone (514) 872-8301, email access to Watson Research Center

George Saon, Daniel Povey, and Geoffrey Zweig

Anatomy of an extremely fast LVCSR decoder