N-GRAM FST INDEXING FOR SPOKEN TERM DETECTION

CHAO LIU
SEP 24, 2012
CONTENTS

1. Introduction to Spoken Term Detection
2. Relate works
   1. N-gram inverted indexing
   2. FST indexing
3. N-gram FST indexing
4. Experiments
5. Conclusion
1. Introduction to Spoken Term Detection
2. Relate works
   1. N-gram inverted indexing
   2. FST indexing
3. N-gram FST indexing
4. Experiments
5. Conclusion
SPEECH TERM DETECTION

- Spoken Term Detection (STD): find all of the occurrences of a specified “term” in a given corpus of speech data. (NIST)
  - Term: a sequence of one or more words. For example: “car”, “New York”.
  - System output: location of the term in audio, a score indicating how likely the term exists.
  - Evaluation: both speed and detection accuracy.

... Kuwait ... car ...
SPEECH TERM DETECTION

We focus on an efficient indexing scheme, which is essentially important for STD on large databases.
CONTENTS

1. Introduction to Spoken Term Detection

2. Relate works
   1. N-gram inverted indexing
   2. FST indexing

3. N-gram FST indexing

4. Experiments

5. Conclusion
**N-GRAM INVERTED INDEXING**

- Get all n-gram fragments with their confidence scores existing in the input lattice, and sort them in chronological order.
- Speed up term searching using inverted list.

![Diagram](https://via.placeholder.com/150)

- **Graph:**
  - Node 1 connected to 2 and 3 with edges labeled a/1.
  - Node 2 connected to 3 with edge labeled b/1.
  - Node 3 connected to 4 with edge labeled a/1.

- **Inverted List:**
  - **a:**
    - 1-2/0.5
    - 1-3/0.5
    - 3-4/0.5
  - **b:**
    - 2-3/0.5
    - 3-4/0.5
  - **a b:**
    - 1-3/0.5
    - 1-4/0.5
  - **b a:**
    - 2-4/0.5
CONTENTS

1. Introduction to Spoken Term Detection
2. Relate works
   1. N-gram inverted indexing
   2. FST indexing
3. N-gram FST indexing
4. Experiments
5. Conclusion
FINITE STATE TRANSDUCER

• Basic parts of FST
  • Input label - phone / n-gram
  • Output label – time period
  • Weight – confidence

• FST operations
  • Determinization
  • Minimization
  • Unification

ac->xz / 6.5
bc->yz / 7.5
FST INDEXING

- Convert lattice to FST by linking initial and final states to all other states.
FST INDEXING

- Compile searching term to FST
- Do composition on term FST and utterance FST
CONTENTS

1. Introduction to Spoken Term Detection
2. Relate works
   1. N-gram inverted indexing
   2. FST indexing
3. N-gram FST indexing
4. Experiments
5. Conclusion
N-GRAM FST INDEXING

- Convert lattice to N-gram fragments
- Compile N-gram fragments to FST
OPTIMIZATION

• Standard operations
  • Determinize, Minimize, RmEpsilon.
  • Viewing it as an acceptor, encoded label (Allauzen and Mohri, 2004)

• Union
  • Corpus

• Terms
FUZZY SEARCH

- OOV words with uncertain pronunciation / mispronounced
  - N-best pronunciation prediction (Wang and King, 2011)

- Just union FSTs together
CONTENTS

1. Introduction to Spoken Term Detection
2. Relate works
   1. N-gram inverted indexing
   2. FST indexing
3. N-gram FST indexing
4. Experiments
5. Conclusion
BACKGROUND

- The ASR system was built with corpora used for train AMI RT05s ASR system
  - 80.2 hours of speech for acoustic model (AM) training
  - 521M words of text for language model (LM) training
  - Phone Error Rate (PER) is 40.49%
  - Average lattice density is 805 nodes / second
- STD Experiments were performed on RT04s and RT05s data sets
  - 489 INV terms and 67 OOV terms for development
  - 255 INV terms and 484 OOV terms for evaluation
EXPERIMENTS

• Metric for accuracy: **Actual Term Weighted Value**
  
  \[
  ATWV = 1 - \text{average} \{ P_{\text{Miss}}(\text{term}) + \beta \cdot P_{\text{FA}}(\text{term}) \}
  \]

• Relevant factors
  
  • N
  
  • Confidence measures
COMPARISON

Searching efficiency

Index size
COMPARISON
## RESULTS ON EVAL SET

<table>
<thead>
<tr>
<th>INV terms</th>
<th>ATWV</th>
<th>Index size/MB</th>
<th>Time/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lattice</td>
<td>0.4782</td>
<td>483</td>
<td>&gt;10^3</td>
</tr>
<tr>
<td>FST indexing</td>
<td>0.4782</td>
<td>959</td>
<td>16.6</td>
</tr>
<tr>
<td>N-gram inverted indexing</td>
<td>0.5310</td>
<td>226</td>
<td>6.0</td>
</tr>
<tr>
<td>N-gram FST indexing</td>
<td>0.5310</td>
<td>943</td>
<td>5.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OOV terms</th>
<th>ATWV</th>
<th>Index size/MB</th>
<th>Time/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lattice</td>
<td>0.2191</td>
<td>483</td>
<td>&gt;10^3</td>
</tr>
<tr>
<td>FST indexing</td>
<td>0.2191</td>
<td>959</td>
<td>19.0</td>
</tr>
<tr>
<td>N-gram inverted indexing</td>
<td>0.2813</td>
<td>226</td>
<td>9.7</td>
</tr>
<tr>
<td>N-gram FST indexing</td>
<td>0.2813</td>
<td>943</td>
<td>6.0</td>
</tr>
<tr>
<td>FST indexing / fuzzy</td>
<td>0.2305</td>
<td>959</td>
<td>81.1</td>
</tr>
<tr>
<td>N-gram inverted indexing / fuzzy</td>
<td>0.3156</td>
<td>226</td>
<td>401.9</td>
</tr>
<tr>
<td>N-gram FST indexing / fuzzy</td>
<td>0.3156</td>
<td>943</td>
<td>30.4</td>
</tr>
</tbody>
</table>
CONTENTS

1. Introduction to Spoken Term Detection
2. Relate works
   1. N-gram inverted indexing
   2. FST indexing
3. N-gram FST indexing
4. Experiments
5. Conclusion
CONCLUSION

- Compared with conventional FST indexing, N-gram FST indexing provides better STD performance by relaxing phone connectivity.
- Compared with the conventional N-gram inverted indexing, this approach is faster and possesses advantages of FSTs in terms of solid theory and rich tools.
- N-gram FST indexing shows significant improvement while doing fuzzy search.
THANK YOU!

Q&A