A KEYWORD-AWARE GRAMMAR FRAMEWORK FOR LVCSR-BASED SPOKEN KEYWORD SEARCH

I-FAN (陳羿帆), GEORGIA TECH, USA
PRESENTER: 吳佳樺
OUTLINE

- INTRODUCTION
- KEYWORD-AWARE GRAMMAR
- REALIZATION OF THE GRAMMAR
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS AND DISCUSSION
- CONCLUSION
OUTLINE

- INTRODUCTION
- KEYWORD-AWARE GRAMMAR
- REALIZATION OF THE GRAMMAR
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS AND DISCUSSION
- CONCLUSION
INTRODUCTION (1/4)

- **Spoken keyword search (KWS)** is a task of detecting a set of preselected keywords in continuous speech.
- Detecting a set of preselected keywords’ applications
  - spoken term detection
  - spoken document indexing and retrieval
  - speech surveillance
- **KWS systems**
  - classic keyword-filler based KWS
  - large vocabulary continuous speech recognition (LVCSR) based KWS.
classic keyword-filler based KWS

- a spoken utterance is represented as a sequence of keywords and nonkeywords
- often achieves a high detection rate using only a small amount of data for acoustic model training
- the systems are restricted to the set of predefined keywords and often produce a great amount of false alarms.

LVCSR based KWS

- LVCSR-based KWS gives better detection results and much fewer false alarms than keywordfiller based KWS
- it often becomes a major performance bottleneck with more detection misses for LVCSR-based KWS in applications where only limited linguistic resources are available.
keyword-aware (KW-aware) grammar framework: integrates the keyword-filler loop grammar into the n-gram LM grammar used by LVCSR-based KWS
reveal that the proposed framework not only preserves the characteristics of high accuracy, low false alarms, and keyword flexibility of LVCSR-based KWS, but also inherits the high detection rate from keyword-filler based KWS under resource-limited conditions.

significant performance improvement for both poorly-trained and well-trained n-gram LMs
we show how to implement the KW-aware grammar with weighted finite-state automata (WFSA) without any approximation.

We also compare the WFSA-realized KW-aware grammar with the previously proposed context-simulated keyword language model (CS-KWLM) interpolated LM approximation to show the similarities and differences between the two realizations.
OUTLINE

- INTRODUCTION
- KEYWORD-AWARE GRAMMAR
- REALIZATION OF THE GRAMMAR
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS AND DISCUSSION
- CONCLUSION
KEYWORD-AWARE GRAMMAR

- When n-gram LMs are poorly trained with limited or topic mismatched data, LVCSR-based KWS usually suffers from high detection misses due to underestimated keyword prior probabilities.

- To alleviate the problem, in the KW-aware grammar framework, probabilities of keywords are boosted by inserting additional standalone keyword paths with appropriate scores to the n-gram LM grammars

\[
P_{KW\text{-}aware} = \max \{ P_{n\text{-}gram}(k|h) \} \\
P_{n\text{-}gram}(k|h) = \prod_{i=0}^{L} P(W_i|h_i)
\]
OUTLINE

- INTRODUCTION
- KEYWORD-AWARE GRAMMAR
- REALIZATION OF THE GRAMMAR
  - Representation of $n$-gram LMs with WFSAs
  - Approximation of the KW-Aware grammar
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS AND DISCUSSION
- CONCLUSION
Representation of \(n\)-gram LMs with WFSAs (1/2)

- In a WFSA representation of an \(n\)-gram LM over the log semiring, each state in the WFSA represents an \(n\)-gram conditioning history \(h_i\).
- Each transitions leaving the state represent a word \(w_i\) with a weight \(-\log(P(w_i|h_i))\) or a backoff transition to a lower-order conditioning history state.
Representation of $n$-gram LMs with WFSAs(2/2)

- A string accepted by the WFSA has a single path through the automaton, and the weight of the string is the sum of the transition weights in that path in a form of negative log probability.

- Given a finite set of state, $Q$, in an $n$-gram WFSA and a string $k = w_1 \ldots w_L$, we denote hist of$[k, Q]$ as the state in $Q$ encoding the conditioning history that matches the end of the string $k$ with the highest order.
Approximation of the KW-Aware grammar

- The boosting effect of is approximated by interpolating the original n-gram LM with a keyword LM.
- The training text of the keyword LM consists of the system keywords prefixed and suffixed by common context terms derived from the original training text.
- This contextsimulated keyword LM (CS-KWLM) has been shown to provide significant performance enhancement for KWS systems.

\[ P_{INT_{LM}}(k|h) = \alpha \times P_{CS-KWLM}(w|h) + (1 - \alpha)P_{LM}(w|h) \]
OUTLINE

- INTRODUCTION
- KEYWORD-AWARE GRAMMAR
- REALIZATION OF THE GRAMMAR
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS AND DISCUSSION
- CONCLUSION
EXPERIMENTAL SETUP (1/2)

- the IARPA Babel OpenKWS13 (Vietnamese) and OpenKWS14 (Tamil) limited language pack (LLP) tasks.
- In both tasks only **10-hour transcribed** audio were used for system training.
- The data are conversational speech between two parties over a telephone channel, which can be landline, cellphone, or phones embedded in vehicles, with the sampling rate set at 8000 Hz.
For both OpenKWS13 and OpenKWS14 systems, the 15-hour evaluation part 1 data (released as evalpart1 by NIST) were used for testing. The evaluation keyword lists contain 4065 and 5576 phrases with out-of-vocabulary words not appearing in the training set for the two tasks respectively. All keyword search systems were LVCSR-based with hybridDNN-HMM acoustic models built with the Kaldi toolkit.
OUTLINE

- INTRODUCTION
- KEYWORD-AWARE GRAMMAR
- REALIZATION OF THE GRAMMAR
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS AND DISCUSSION
- CONCLUSION
EXPERIMENTAL RESULTS AND DISCUSSION (1/4)

Table I. Vietnamese grammar WFSAs used in the three systems

<table>
<thead>
<tr>
<th>Vietnamese grammars</th>
<th># arcs</th>
<th># states</th>
<th>File size</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-gram baseline</td>
<td>38,713</td>
<td>17,616</td>
<td>812 Kb</td>
</tr>
<tr>
<td>KW-aware grammar (global $\kappa=0.00005$)</td>
<td>66,913</td>
<td>24,215</td>
<td>1.3 Mb</td>
</tr>
<tr>
<td>CS-KWLM Int ($\alpha=0.6$)</td>
<td>381,461</td>
<td>165,063</td>
<td>7.8 Mb</td>
</tr>
</tbody>
</table>
### Table II. Performance on the Vietnamese LLP evalpart1 data.

<table>
<thead>
<tr>
<th>Vietnamese [evalpart1]</th>
<th># Miss</th>
<th>ATWV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$-gram baseline</td>
<td>2562</td>
<td>0.2098</td>
</tr>
<tr>
<td>KW-aware framework</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KW-aware grammar (global $\kappa=0.00005$)</td>
<td>1589</td>
<td>0.3224</td>
</tr>
<tr>
<td>CS-KWLM Int ($\alpha=0.6$)</td>
<td>1651</td>
<td>0.3287</td>
</tr>
</tbody>
</table>
### Table III. Performance on the Tamil LLP evalpart1 data.

<table>
<thead>
<tr>
<th>Tamil [evalpart1]</th>
<th># Miss</th>
<th>ATWV</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-gram baseline</td>
<td>3663</td>
<td>0.2128</td>
</tr>
<tr>
<td>KW-aware grammar (global $\kappa=0.0000347$)</td>
<td>2830</td>
<td>0.3102</td>
</tr>
<tr>
<td>CS-KWLM Int ($\alpha=0.3$)</td>
<td>2689</td>
<td>0.3160</td>
</tr>
</tbody>
</table>
EXPERIMENTAL RESULTS AND DISCUSSION (4/4)
OUTLINE

- INTRODUCTION
- KEYWORD-AWARE GRAMMAR
- REALIZATION OF THE GRAMMAR
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS AND DISCUSSION
- CONCLUSION
CONCLUSION

- In this paper, we proposed an exact realization, which was a missing block in the current KW-aware framework, of the KW-aware grammar.

- Experimental results on Babel Vietnamese and Tamil LLP tasks show the exact realization was very compact and outperformed our previous approximation method for long keywords; while for short keywords, the performances of the two systems were similar.

- We also showed that the significant performance improvement of the proposed KW-aware framework over the n-gram baseline is consistent across languages.