LOW-RESOURCE KEYWORD SEARCH STRATEGIES FOR TAMIL

Nancy F. Chen, TECHNOLOGY AND RESEARCH (A*STAR), Singapore.

presenter: 吳佳樺
OUTLINE

• INTRODUCTION
• RELATION TO PRIOR WORK
• LOW-RESOURCE KEYWORD SEARCH STRATEGIES
• EXPERIMENTS
• DISCUSSION
INTRODUCTION

RELATION TO PRIOR WORK

LOW-RESOURCE KEYWORD SEARCH STRATEGIES

EXPERIMENTS

DISCUSSION
INTRODUCTION

• Keyword search (KWS) is a detection task where the goal is to find all occurrences of an orthographic term (e.g., word or phrase) from audio recordings.

• Applications of KWS include spoken document indexing and retrieval and spoken language understanding.

• KWS systems:
  1. classic keyword-filler based KWS.
  2. large vocabulary continuous speech recognition (LVCSR) based KWS
KWS systems

classic keyword-filler based KWS

• a spoken utterance is represented as a sequence of keywords and non-keywords.
• Keyword-filler based systems often achieve high detection rate using only a small dataset for acoustic model training
• but they do not scale well when the number of keywords increases.

(LVCSR) based KWS

• flexible in handling a large number of keywords.
• yet require sufficiently large amounts of transcribed training data to achieve good performance.
• LVCSR-based KWS has worked well on resource-rich languages like English.
keyword search system for low-resource languages
INTRODUCTION

RELATION TO PRIOR WORK
• Active Learning for Selecting Audio to Transcribe
• Keyword-Aware Language Modeling
• Subword Modeling: Morphemes and Homophones

LOW-RESOURCE KEYWORD SEARCH STRATEGIES

EXPERIMENTS

DISCUSSION
Active Learning for Selecting Audio to Transcribe

- Transcribing speech data is **time-consuming and labor-intensive**, especially for low resource languages where linguistic expertise is limited or lacking.
- In prior work, most approaches select utterances in a **greedy fashion** according to their **utility scores** (e.g., confidence scores from automatic speech recognition (ASR)).
- Similar to the confidence-based approaches, use **entropy reduction** to select unlabeled utterances.
Keyword-Aware Language Modeling

• If keyword queries are known a priori, one can leverage such knowledge to improve KWS performance.
• Previous work has shown how to exploit keyword information for acoustic modeling and decoding.
• Our previous efforts exploit keyword information in language modeling in Vietnamese.
• In this work, we investigate our approach in Tamil and further expand it to a framework integrating advantages from both keyword-filler based KWS and LVCSR-based KWS.
Subword Modeling: Morphemes and Homophones

- Mainstream LVCSR systems suffer from out-of-vocabulary (OOV) issues.
- For morphologically-rich languages like Tamil, OOV rate is especially high.
- While phones are commonly used to help resolve OOVs, morphs (automatically parsed morphemes) have also been used in ASR.
- In this work, we mitigate the data sparsity issue of the morphologically-rich vocabulary in Tamil by integrating morphs in the lexicon and language models.
OUTLINE

• INTRODUCTION
• RELATION TO PRIOR WORK
• LOW-RESOURCE KEYWORD SEARCH STRATEGIES
  • Submodular Optimization to Select Audio to Transcribe
  • Keyword-Aware Language Modeling (LM)
  • Word-Morph Interpolated Language Model
• EXPERIMENTS
• DISCUSSION
Submodular Optimization to Select Audio to Transcribe

\[ \overline{m}_u(S) = \frac{m_u(S)}{\sum_{u \in U} m_u(S)} \]

\[ D_{KL}(P||\overline{m}_u(S)) = \sum_{u \in U} p_u \log p_u - \sum_{u \in U} p_u \log(m_u(S)) + \log(\sum_{u \in U} m_u(S)) \]

\[ = \text{const.} + \log(\sum_{u \in U} m_u(S)) - \sum_{u \in U} p_u \log(m_u(S)) \]

\[ f(S) = \log(\sum_{u \in U} m_u(S)) - D_{KL}(P||\overline{m}_u(S)) \]

\[ = \sum_{u \in U} p_u \log(m_u(S)). \]
Keyword-Aware Language Modeling (LM)

- Let \( q = (w_1, w_2, \ldots, w_L) \) be an \( L \)-word query. In keyword-filler based KWS, the prior probability \( P(q) \) is by default set to \( P(q) = 1/N \), often resulting in high false alarms. (Typically \( N \approx 100 \).)

\[
P_{\text{LVCSR}}(q) = \sum_{h \in H} P(q|h) \approx \sum_{h \in H} \prod_{i=1}^{L} P_{n-\text{gram}}(w_i|h_i(h, q))
\]

\[
P_{\text{KW-aware}}(q) = \max\{P_{\text{LVCSR}}(q), k(q)\}
\]
Word-Morph Interpolated Language Model

- Representing out-of-vocabulary (OOV) entries using morphs (automatically parsed morphemes) is insufficient to resolve data sparsity issues with morphologically-rich languages like Tamil.
- If the morphbased lexical entries have low occurrences, the miss probability of such keywords are still high despite it no longer being OOV.
- exploit word-morph interpolated language models (LM) to provide smoother estimates:

\[ \lambda_{W-M} = \alpha \lambda_W + \beta \lambda_M + (1 - \alpha - \beta) \lambda_H. \]
OUTLINE

• INTRODUCTION
• RELATION TO PRIOR WORK
• LOW-RESOURCE KEYWORD SEARCH STRATEGIES
• EXPERIMENTS
• DISCUSSION
EXPERIMENTS

- the baseline ATWV results when using word 3-gram LM for Limited Language Pack (LLP) and Full Language Pack (FLP).

<table>
<thead>
<tr>
<th>Transcription Condition</th>
<th>Baseline ATWV</th>
<th>Keyword-Aware LM ATWV</th>
<th>Relative Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLP: 10 hr</td>
<td>0.2313</td>
<td>0.3182</td>
<td>37.6%</td>
</tr>
<tr>
<td>FLP: 60 hr</td>
<td>0.4222</td>
<td>0.4852</td>
<td>14.9%</td>
</tr>
</tbody>
</table>

Table 2. Word-Morph LM outperforms word LM.

<table>
<thead>
<tr>
<th>Transcription Condition</th>
<th>Word LM ATWV</th>
<th>Word-Morph LM ATWV</th>
<th>Rel. Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLP: 10 hr</td>
<td>0.2313</td>
<td>0.2418</td>
<td>4.5</td>
</tr>
<tr>
<td>FLP: 60 hr</td>
<td>0.4222</td>
<td>0.4363</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Actual term-weighted value (ATWV)
## EXPERIMENTS

<table>
<thead>
<tr>
<th>Transcription Condition</th>
<th>ATWV</th>
<th>OOV counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline-1: Random 10 hr subset</td>
<td>0.2386</td>
<td>1171</td>
</tr>
<tr>
<td>Baseline-2: NIST-LLP (10 hr subset)</td>
<td>0.2474</td>
<td>1686</td>
</tr>
<tr>
<td>Proposed submodular 10 hr subset</td>
<td>0.2857</td>
<td>972</td>
</tr>
<tr>
<td>Upper bound: NIST-FLP (full 60 hr)</td>
<td>0.4363</td>
<td>407</td>
</tr>
</tbody>
</table>
OUTLINE

• INTRODUCTION
• RELATION TO PRIOR WORK
• LOW-RESOURCE KEYWORD SEARCH STRATEGIES
• EXPERIMENTS
• DISCUSSION
DISCUSSION

• investigated strategies for low-resource keyword search
  • Submodular optimization for selecting data to transcribe: to generalize well in languages other than Tamil since similar approaches works in Mandarin LVCSR.
  • subword modeling of morphemes and homophones.
  • keyword-aware language modeling: works for Vietnamese.
• While our LVCSR-KWS work in Tamil and Vietnamese [24] focus on text queries, we have inspired strategies used in spoken term detection of audio queries.