Spoken Language Translation Graphs Re-decoding using Automatic Quality Assessment
Laurent Besacier, Benjamin Lecouteux, Ngoc Luong, Ngoc Le

To cite this version:
Laurent Besacier, Benjamin Lecouteux, Ngoc Luong, Ngoc Le. Spoken Language Translation Graphs Re-decoding using Automatic Quality Assessment. ASRU, 2015, Scotsdale, United States. hal-02095256

HAL Id: hal-02095256
https://hal.archives-ouvertes.fr/hal-02095256
Submitted on 10 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract
This paper investigates how automatic quality assessment of spoken language translation (SLT) can help re-decoding SLT output graphs and improving the overall speech translation performance. Using robust word confidence measures (from both ASR and MT) to re-decode the SLT graph leads to a significant BLEU improvement (more than 2 points) compared to our SLT baseline (French-English task).

Introduction

- **Context**
  - Automatic quality assessment of spoken language translation (SLT)
  - Also named confidence estimation (CE)
  - Pointing out correct parts and errors in a speech translated output
- **Useful for**
  - Interactive speech to speech translation
  - Computer- assisted translation (from speech or text)
- **Claim**
  - An accurate CE can also help to improve SLT itself through search graph re-decoding

Formalisation

\[ x_t \text{ source signal}, f - (f_1, f_2, ..., f_N) \text{ transcription of } x_t. \]
\[ \hat{e} = (e_1, e_2, ..., e_N) \text{ translation of } f \text{ and } \hat{e} = \text{argmax}(p(e|x_t, f)) \]

Word Confidence Estimation (WCE) can be seen as finding sequence \( q \) where \[ q = (q_1, q_2, ..., q_N) \text{ and } q_i \in \{ \text{good, bad} \} \]
\[ \hat{e} = \text{argmax} \sum_{e,q} p(e,q|x_t, f) \hat{f} = \text{argmax} \sum_{e,p,q} p(q|x_t, f, e) \cdot p(e|x_t, f) \]
\[ \hat{e} = \text{argmax} \sum_{p,q|x_t, f, e} \cdot p(e|x_t, f) \]
\[ p(q|x_t, f, e) : \text{WCE component} \]
\[ p(e|x_t, f) : \text{SLT component} \]

SLT Search Graph (SG) Re-decoding

For SLT N-best hypotheses \( e^N = (e_1^N, e_2^N, ..., e_N^N) \), each \( j \)-th word in the hypothesis \( e_j \), denoted by \( e_j \), has a quality label, \( q_j \). For all hypothesis \( H_k \) in \( e_j \), the new transition cost is defined by:

\[
\text{transition}(H_k) = \text{transition}(H_k) + \text{reward}(q_j) \quad \text{if } q_j \text{ = good} \\
\text{otherwise} \]

\[
\text{with} \\
\text{penalty}(q_j) = -\text{reward}(q_j) - \beta \cdot \frac{\text{score}(H_k)}{\# \text{words}(H_k)}
\]

Analysis of SLT Hypotheses

### Example 1
- hat de démission des employés peut déboucher sur une démission mortelle
- **1st word**
  - \( e_1^N \)
  - paternité
  - **re-decoding**
- **2nd word**
  - \( e_2^N \)
  - démission
  - **re-decoding**
- **3rd word**
  - \( e_3^N \)
  - mortelle
  - **re-decoding**

### Example 2
- cela a un impact sur l’intervention d’abord parfaitement bien décrite avec une intervention post-operative voulue normale
- **1st word**
  - \( e_1^N \)
  - impact
  - **re-decoding**
- **2nd word**
  - \( e_2^N \)
  - sur
  - **re-decoding**
- **3rd word**
  - \( e_3^N \)
  - parfaitement
  - **re-decoding**

### Example 3
- general motors repousse jusqu’en janvier le plan pour ouvrir
- **1st word**
  - \( e_1^N \)
  - motors
  - **re-decoding**
- **2nd word**
  - \( e_2^N \)
  - repousse
  - **re-decoding**
- **3rd word**
  - \( e_3^N \)
  - jusqu’en
  - **re-decoding**

<table>
<thead>
<tr>
<th>Example</th>
<th>Transition</th>
<th>Re-decoding</th>
<th>WCE</th>
<th>SLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st word</td>
<td>e_1^N</td>
<td>paternité</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd word</td>
<td>e_2^N</td>
<td>démission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd word</td>
<td>e_3^N</td>
<td>mortelle</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table : Examples of French SLT output with and w/o re-decoding

Building an Efficient WCE System

\[
\hat{q} = \text{argmax} \sum_{e,q} p(q|x_t, f, e) \\
\text{need training data with quadruplet } (x_t, f, e, q) \text{ available, so instead we compute: } \\
\hat{q} = \text{argmax} \sum_{q} p_{\text{ASR}}(q|x_t, f) \cdot \text{part}(q, e, f) \\
\]

- **p_{\text{ASR}}(q|x_t, f)**
  - system described in [1]
  - acoustic / graph / linguistic / lexical features
  - boosting classifier
- **part(q, e, f)**
  - system described in [2]
  - multiple features and CRF classifier
  - using our open-source toolkit available online

\[ \text{http://github.com/besacier/WCE-LIG} \]

Experimental Setting

- **French ASR**
  - Kaldi toolkit [3]
  - HMM/SGMM / 3-grams
- **French-English SMT**
  - Moses toolkit [5]
  - 1.6M parallel sent.
  - 48M monolingual sent.
  - medium-size system
  - WMT shared task

<table>
<thead>
<tr>
<th>Test corpus</th>
<th>French SMT</th>
<th>French-English SMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2643 French utterances</td>
<td>5-h of speech</td>
<td>1.6M parallel sent.</td>
</tr>
<tr>
<td>5-h of speech</td>
<td>48M monolingual sent.</td>
<td>medium-size system</td>
</tr>
<tr>
<td>Cross-validation for tuning / decoding</td>
<td>WMT shared task</td>
<td></td>
</tr>
<tr>
<td>Quality labels q \in { \text{good, bad} }</td>
<td>obtained with TERP-A toolkit [4]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>