TALP: Xgram-based Spoken Language Translation System

Adrià de Gispert
José B. Mariño
Outline

- Overview
- Translation generation
- Training
- IWSLT'04 Chinese-English supplied task results
- Conclusion and further work
Overview

- TALP Statistical Machine Translation (SMT)
- Integrated speech-text approach
- Finite-State Transducer (FST) implementation
- Automatically learnt from parallel corpus
- Bilingual units called tuples
Translation generation

- Maximising joint probability
- Variable-length N-gram of bilingual units (tuples)

\[
\hat{f} = \arg\max_{f} p(e, f)
\]

\[
p(e, f) = \prod_{n=1}^{N} p((e, f)_n | (e, f)_{n-1} \ldots (e, f)_{n-X+1})
\]

\[
(e, f)_n = (e_{i(n)} \ldots e_{i(n)+I(n)}, f_{j(n)} \ldots f_{j(n)+J(n)})
\]
FST implementation

- Search for best-scoring path
- Speech translation: include acoustic models
Training

Parallel Corpus

Pre-processing

- Automatic from parallel corpus
- Currently based on Giza++ alignments

Word alignment

Tuples extraction

Embedded words dict

X-gram estimation
Preprocessing

- Particular for each pair of languages
- Categorisation: personal names, dates, times, numbers, ...

**Chinese-English** IWSLT'04 supplied track:
- Clearing out punctuation: no gain
- Segmentation of longer sentences

<table>
<thead>
<tr>
<th></th>
<th># sentences</th>
<th>Lavg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>20 K (22.2 K)</td>
<td>9.1 (8.2)</td>
</tr>
<tr>
<td>English</td>
<td>9.4 (8.5)</td>
<td></td>
</tr>
</tbody>
</table>
Word alignment

- Standard GIZA++ alignments \( 1^5 H^5 3^3 4^3 \)
  - Source-to-Target (s2t)
  - Target-to-Source (t2s)
  - Union
  - Intersection

- Tuples can be extracted from any alignment
- Usually **union** and **s2t** are used
Tuples extraction

- Tuples are bilingual units containing
  - one or more source words
  - zero, one or more target words

- Subset of phrases, unique under following conditions

- Example

1. Monotonous segmentation of the pair
2. Words are consecutive along both source and target
3. No word in the tuple is aligned to a word outside the tuple
4. The tuple cannot be decomposed without violating 1-3
Embedded words dictionary

- Embedded words
  - translation appears always inside a tuple (never isolated)
  - an 'accurate' dictionary is built
- New unigrams (history independent)

**DICTIONARY ENTRIES:**

Given a source word, look for the most freq. aligned words

1. Target words are consecutive
2. Target words are aligned *only* to the source word
$ (f_1 f_2, e_1) (f_3, e_2 e_3 e_4) (f_4 f_5 f_6, e_5 e_6) $ 

- Usually, maximum memory is 3
- Pruning strategies
  - Min. number of times a certain history must occur $ k $
  - Threshold of divergence between output prob. distributions for two nodes sharing recent history $ f $

$k$ is untouched ($= 1$)  
$f$ used for slight pruning ($f = 0.2$)
## Chinese-English supplied track

### Results on development set

- Union (aU) vs. s2t alignment (a2)
- Normal vs. segmented corpus (seg)
- Normal vs. FST pruning (f)

### Table of Results

<table>
<thead>
<tr>
<th>run</th>
<th>BLEU</th>
<th>NIST</th>
<th>WER</th>
<th>PER</th>
<th>GTM</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>aU</td>
<td>0.244</td>
<td>5.169</td>
<td>0.615</td>
<td>0.529</td>
<td>0.591</td>
<td>7</td>
</tr>
<tr>
<td>aU,seg</td>
<td>0.251</td>
<td>5.187</td>
<td>0.607</td>
<td>0.521</td>
<td><strong>0.595</strong></td>
<td>7</td>
</tr>
<tr>
<td>aU,seg,f</td>
<td>0.255</td>
<td><strong>5.210</strong></td>
<td><strong>0.603</strong></td>
<td><strong>0.518</strong></td>
<td>0.594</td>
<td>7</td>
</tr>
<tr>
<td>a2</td>
<td><strong>0.319</strong></td>
<td>3.789</td>
<td>0.614</td>
<td>0.552</td>
<td>0.573</td>
<td>16</td>
</tr>
<tr>
<td>a2,seg</td>
<td>0.318</td>
<td>3.871</td>
<td>0.606</td>
<td>0.546</td>
<td>0.573</td>
<td>18</td>
</tr>
<tr>
<td>a2,seg,f</td>
<td>0.314</td>
<td>3.678</td>
<td>0.607</td>
<td>0.548</td>
<td>0.570</td>
<td>19</td>
</tr>
<tr>
<td>no embed. dict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aU,seg-D</td>
<td>0.264</td>
<td>4.741</td>
<td>0.606</td>
<td>0.524</td>
<td>0.592</td>
<td>7</td>
</tr>
<tr>
<td>a2,seg-D</td>
<td>0.315</td>
<td>3.706</td>
<td>0.607</td>
<td>0.547</td>
<td>0.571</td>
<td>19</td>
</tr>
</tbody>
</table>

### Runs

- run A
- run B
Automatic evaluation results

- Statistics of submitted runs

<table>
<thead>
<tr>
<th>run</th>
<th>tuples</th>
<th>vcb</th>
<th>length</th>
<th>embed</th>
</tr>
</thead>
<tbody>
<tr>
<td>aU,seg,f</td>
<td>97 K</td>
<td>27 K</td>
<td>3.9</td>
<td>4.7 K</td>
</tr>
<tr>
<td>a2,seg,f</td>
<td>140 K</td>
<td>29 K</td>
<td>2.9</td>
<td>1.5 K</td>
</tr>
</tbody>
</table>

- Longer tuples with Union (more embedded)
- Many tuples to NULL with s2t (28% over total, 7.5% union)

- Results

<table>
<thead>
<tr>
<th>run</th>
<th>BLEU</th>
<th>NIST</th>
<th>WER</th>
<th>PER</th>
<th>GTM</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>aU,seg,f</td>
<td>0.279</td>
<td>6.778</td>
<td>0.556</td>
<td>0.465</td>
<td>0.647</td>
<td>5</td>
</tr>
<tr>
<td>a2,seg,f</td>
<td>0.331</td>
<td>5.391</td>
<td>0.550</td>
<td>0.490</td>
<td>0.620</td>
<td>11</td>
</tr>
</tbody>
</table>

- Contrast BLEU vs. NIST (related to length)
Manual evaluation results

- Results

<table>
<thead>
<tr>
<th>run</th>
<th>fluency</th>
<th>adequacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>aU,seg,f</td>
<td>2.792</td>
<td>3.022</td>
</tr>
</tbody>
</table>

- Expected fluency deficiency  no explicit long reordering
- 'Much of the information' is transmitted

Examples:

Translation:  that what time start
Reference1:  what time does it start

Translation:  stomach very hurts
Reference1:  i have a severe pain in my stomach
Conclusion and further research

- Tuple-based FST translation system presented
- Adequate for pairs of languages *similar in word-order*

- Further research
  - Embedded N-grams
  - Generalization of tuples
  - Explicit reordering techniques
Thanks for attention

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
TALP Research Center
Universitat Politècnica de Catalunya (UPC)
Barcelona