Machine Translation: Examples

Atlanta, preso il killer del palazzo di Giustizia

Atlanta, taken the killer of the palace of Justice

Levels of Transfer

General Approaches

The Coding View

MT System Components
Language Models

- Language Models
  - Any probabilistic model capable of assigning probabilities to sentences
  - Usually n-gram models, but also PCFGs
  - Exact same technology (and software) as in ASR
  - Train on a huge collection of monolingual corpora (documents in the target language)

Parallel Corpora

- Parallel corpora (or bitexts)
  - Collection of source-target translation pairs
  - Main resource for learning a translation model
  - Either naturally occurring (e.g. parliamentary proceedings, news translation services) or commissioned

Building a Translation Model

- Steps in building a simple statistical translation model
  - Match up words in training sentence pairs (word alignment)
  - Learn a lexicon from these alignments
  - Learn larger phrases

1-to-Many Alignments

- The HMM Alignment Model
  - The HMM model (Vogel 96)
  - Re-estimate using the forward-backward algorithm
  - Handling nulls requires some care
  - Note: alignments are not provided, but induced
**Examples: Translation and Fertility**

<table>
<thead>
<tr>
<th>the</th>
<th>not</th>
</tr>
</thead>
<tbody>
<tr>
<td>le</td>
<td>0.207</td>
</tr>
<tr>
<td>les</td>
<td>0.155</td>
</tr>
<tr>
<td>l'</td>
<td>0.096</td>
</tr>
<tr>
<td>ce</td>
<td>0.018</td>
</tr>
</tbody>
</table>

**Phrases vs Word Models**

<table>
<thead>
<tr>
<th>he is nodding</th>
</tr>
</thead>
<tbody>
<tr>
<td>il hoche la tête</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Examples:</th>
<th>Translation and Fertility</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>not</td>
</tr>
<tr>
<td>le</td>
<td>0.207</td>
</tr>
<tr>
<td>les</td>
<td>0.155</td>
</tr>
<tr>
<td>l'</td>
<td>0.096</td>
</tr>
<tr>
<td>ce</td>
<td>0.018</td>
</tr>
</tbody>
</table>

**Extracting Phrases**

**Basic Phrase-Based Model**

- Now we have a phrase table:
  - A huge list of translation phrases (e.g. 1M phrases)
  - Each phrase has a probability $P(\text{f|e})$

- When we see a new input sentence:
  - Grow a translation left to right
  - Extend translation using known phrases
  - Also multiply by language model score

**Decoding**

- Probabilities at each step include LM and TM
Madame la présidente, votre présidence de cette institution a été marquante.

Mrs Fontaine, your presidency of this institution has been outstanding.

Madam President, president of this house has been discoveries.

Madam President, your presidency of this institution has been impressive.

Je vais maintenant m’exprimer brièvement en irlandais.

I shall now speak briefly in Irish.

I will now speak briefly in Ireland.

Nous trouvons en vous un président tel que nous le souhaitions.

We think that you are the type of president that we want.

We are in you a president as the wanted.

We are in you a president as we the wanted.

**Some Output**

**Translations**

- Even human translators aren’t perfect:
  - In an Austrian ski hotel:
    Not to perambulate the corridors in the hours of repose in the boots of ascension.
  - In a Copenhagen airline ticket office:
    We take your bags and send them in all directions.
  - From a brochure of a car rental firm in Tokyo: When passenger of foot heave in sight, tootle the horn. Trumpet him melodiously at first, but if he still obstacles your passage then tootle him with vigor.

http://www.englishfirst.org/13166/funnytranslations.htm