Machine Translation: Examples

Atlanta, preso il killer del palazzo di Giustizia

ATLANTA - La grande paura che per 20 ore ha attanagliato Atlanta è finita. Brian B nợ, l'uomo che aveva ucciso tre persone a palazzo di Giustizia e che era stato allontanato dalle polizei, è stato consegnato alla polizia, dopo aver cercato rifugio nell'alloggio di una donna in un complesso di appartamenti alla periferia della città. Per tutta Atlanta, il centro della città, 1500 mila persone e 400000 di (150 000), sono di una popolosaarea metropolitana, era rimasto paralizzata.

Atlanta, taken the killer of the palace of Justice

ATLANTA - The great fear that for 20 hours has gripped Atlanta is ended. Brian Nichols, the man who killed three people in a courthouse and who had been removed from the police, is in the hands of the police, after he tried to hide in the room of a woman in a complex of apartments on the edge of the city. For all of Atlanta, the city center of the city, 1500000 people and 4000000 of (1500000), were of a populousmetropolitan area, was still paralyzed.

Word-Level: Examples

la politique de la haine .
the policy of the hatred .

nous avons signé le protocole .
we did sign the memorandum of agreement .

où était le plan solide ?
where was the solid plan ?

where was the economic base ?

(forward original) (reference translation)

IBM4+ (to grams+stack)

Phrasal MT: Examples

Le président vénitien Berwick Macke a annoncé vendredi sa volonté de réformer le système du tourisme de l'hôtel de ville de Venise. Les chiffres sur les touristes accueillis par le centre de tourisme de l'hôtel de ville montrent que le tourisme est au plus bas depuis plusieurs années. "Il est temps de réfléchir sur la restructuration du tourisme et sa gestion de manière plus efficace," a déclaré le préfet. "Pour les touristes il est temps de réfléchir sur la restructuration et sa gestion de manière plus efficace,"

General Motors and Chrysler are engaged in negotiations with the union to end the strike without any new wages. "We must make sacrifices and be flexible," they warned. "Everyone should get behind the strike and agree to a peaceful restructuring,"

U.S. President Barack Obama is expected Monday to announce his administration's plan to prevent new bankruptcy for U.S. billion dollars, and stabilize the financial system. January is a restructuring plan based on a total of 12 billion dollars in additional aid...

Interim CEO Scott, the president has clearly stated the government does not back more without strong negotiations. "We must make sacrifices and be flexible," they warned. "Everyone should get behind the strike and agree to a peaceful restructuring..."
Metrics

MT: Evaluation

- Human evaluations: subject measures, fluency/adequacy
- Automatic measures: n-gram match to references
  - NIST measure: n-gram recall (worked poorly)
  - BLEU: n-gram precision (no one really likes it, but everyone uses it)
  - Lots more: TER, TTER, METEOR...  
  - BLEU:
    - \( P_1 = \) unigram precision
    - \( P_2, P_3, P_4 = \) bi-, tri-, 4-gram precision
    - Weighted geometric mean of \( P_1 - P_4 \)
    - Brevity penalty (why?)
    - Somewhat hard to game...
    - Magnitude only meaningful on same language, corpus, number of references, probably only within system types...

Automatic Metrics Work (?)

slide from G. Doddington (NIST)

Systems Overview

Corpus-Based MT

Modeling correspondences between languages

Sentence-aligned parallel corpus:

- Yo lo haré mañana
  - I will do it tomorrow
- Hasta pronto
  - See you soon
- Hasta pronto
  - See you around

Machine translation system:

- Yo lo haré pronto
  - I will do it soon
- See you tomorrow

Phrase-Based System Overview

- Sentence-aligned corpus
- Word alignments
- Phrase table (translation model)

Many slides and examples from Philipp Koehn or John Dellers
Word Alignment

What is the anticipated cost of collecting fees under the new proposal?

En vertu des nouvelles propositions, quel est le coût prévu de perception des droits?

Unsupervised Word Alignment

- Input: a bitext: pairs of translated sentences
- Output: alignments: pairs of translated words
  - When words have unique sources, can represent as a forward alignment function a from French to English positions

1-to-Many Alignments

1-to-Many Alignments

- Method 1: use in an end-to-end translation system
  - Hard to measure translation quality
  - Option: human judges
  - Option: reference translations (NIST, BLEU)
  - Option: combinations (HTER)
  - Actually, no one uses word-to-word models alone as TMs

- Method 2: measure quality of the alignments produced
  - Easy to measure
  - Hard to know what the gold alignments should be
  - Often does not correlate well with translation quality (like perplexity in LMs)
Alignment Error Rate

- **Alignment Error Rate**

  \[ \text{AER}(A, S, P) = \left(1 - \frac{|A \cap S| + |A \cap P|}{|A| + |S|}\right) \]

  \[= \left(1 - \frac{3 + 3}{3 + 4}\right) = \frac{1}{7} \]

IBM Model 1: Allocation

IBM Model 1 (Brown 93)

- Alignments: a hidden vector called an alignment specifies which English source is responsible for each French target word.

\[ e = \alpha_1 \ldots \alpha_j \]

\[ P(f, u|e) = \prod_{j} P(\alpha_j = 0) P(f_j|\alpha_j) = \prod_{j} \frac{1}{j+1} P(f_j|\alpha_j) \]

\[ P(f|e) = \sum_{A} P(f, A|e) \]

IBM Models 1/2

Problems with Model 1

- There’s a reason they designed models 2-5!

- Problems: alignments jump around, align everything to rare words.

- Experimental setup:
  - Training data: 1.1M sentences of French-English text, Canadian Hansards.
  - Evaluation metric: alignment error rate (AER).
  - Evaluation data: 447 hand-aligned sentences

Intersected Model 1

- Post-intersection: standard practice to train models in each direction then intersect their predictions [Och and Ney, 03]

- Second model is basically a filter on the first
  - Precision jumps, recall drops
  - End up not guessing hard alignments

<table>
<thead>
<tr>
<th>Model</th>
<th>P/R</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 E&gt;F</td>
<td>82/58</td>
<td>30.6</td>
</tr>
<tr>
<td>Model 1 F&gt;E</td>
<td>85/56</td>
<td>28.7</td>
</tr>
<tr>
<td>Model 1 AND</td>
<td>96/40</td>
<td>34.8</td>
</tr>
</tbody>
</table>
Joint Training?

- Overall:
  - Similar high precision to post-intersection
  - But recall is much higher
  - More confident about positig non-null alignments

<table>
<thead>
<tr>
<th>Model</th>
<th>P/R</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 E→F</td>
<td>82/58</td>
<td>30.6</td>
</tr>
<tr>
<td>Model 1 F→E</td>
<td>85/58</td>
<td>28.7</td>
</tr>
<tr>
<td>Model 1 AND</td>
<td>96/46</td>
<td>34.8</td>
</tr>
<tr>
<td>Model 1 INT</td>
<td>93/69</td>
<td>19.5</td>
</tr>
</tbody>
</table>

IBM Model 2: Global Monotonicity

Monotonic Translation

Japan shaken by two new quakes
Le Japon secoué par deux nouveaux séismes

Local Order Change

Japan is at the junction of four tectonic plates
Le Japon est au confluent de quatre plaques tectoniques

IBM Model 2

- Alignments tend to the diagonal (broadly at least)

\[
P(f, c|e) = \prod_j P(a_j = i, j, f) P(f_j|e_i) \\
1^{d(e_i - i - j | f)} \\
\frac{1}{2^d - \delta(i, j)}
\]

- Other schemes for biasing alignments towards the diagonal:
  - Relative vs absolute alignment
  - Asymmetric distances
  - Learning a full multinomial over distances

EM for Models 1/2

- Model 1 Parameters:
  - Translation probabilities (1+2): \( P(f_j|e_i) \)
  - Distortion parameters (2 only): \( P(a_j = i, j, f) \)

- Start with \( P(f_j|e_i) \) uniform, including \( P(f_j|null) \)
- For each sentence:
  - For each French position \( j \)
    - Calculate posterior over English positions
    - Increment count of word \( f \) with word \( e_i \) by these amounts
- (or just use best single alignment)
- Iterate until convergence
Example

HMM Model: Local Monotonicity

Phrase Movement

The HMM Model

The HMM Model

- Model 2 preferred global monotonicity
- We want local monotonicity:
  - Most jumps are small
- HMM model (Vogel 96)
  \[
  P(l, w|x) = \prod_j P(u_j|x, w_j)P(f_j|x)
  \]
  - Re-estimate using the forward-backward algorithm
  - Handling nulls requires some care
- What are we still missing?

HMM Examples
AER for HMMs

<table>
<thead>
<tr>
<th>Model</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 INT</td>
<td>19.5</td>
</tr>
<tr>
<td>HMM E→F</td>
<td>11.4</td>
</tr>
<tr>
<td>HMM F→E</td>
<td>10.8</td>
</tr>
<tr>
<td>HMM AND</td>
<td>7.1</td>
</tr>
<tr>
<td>HMM INT</td>
<td>4.7</td>
</tr>
<tr>
<td>GIZA M4 AND</td>
<td>6.9</td>
</tr>
</tbody>
</table>

Models 3, 4, and 5: Fertility

IBM Models 3/4/5

Example: Idioms

Example: Morphology
### Some Results

[Och and Ney 03]

<table>
<thead>
<tr>
<th>Model</th>
<th>Training scheme</th>
<th>0.5K</th>
<th>8K</th>
<th>128K</th>
<th>1-47M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dice</td>
<td></td>
<td>30.9</td>
<td>43.4</td>
<td>39.6</td>
<td>36.9</td>
</tr>
<tr>
<td>Dice-C</td>
<td></td>
<td>40.3</td>
<td>37.6</td>
<td>35.0</td>
<td>34.0</td>
</tr>
<tr>
<td>Model 1</td>
<td>$i^1$</td>
<td>40.6</td>
<td>33.6</td>
<td>28.6</td>
<td>25.9</td>
</tr>
<tr>
<td>Model 2</td>
<td>$i^2$</td>
<td>46.7</td>
<td>29.3</td>
<td>22.0</td>
<td>19.5</td>
</tr>
<tr>
<td>HMM</td>
<td>$i^2$</td>
<td>26.3</td>
<td>23.3</td>
<td>15.0</td>
<td>10.8</td>
</tr>
<tr>
<td>Model 3</td>
<td>$i^2i^3$</td>
<td>43.6</td>
<td>27.5</td>
<td>20.5</td>
<td>18.0</td>
</tr>
<tr>
<td>Model 4</td>
<td>$i^3i^4$</td>
<td>27.5</td>
<td>22.5</td>
<td>16.6</td>
<td>13.2</td>
</tr>
<tr>
<td>Model 5</td>
<td>$i^2i^3i^4$</td>
<td>41.7</td>
<td>25.1</td>
<td>17.3</td>
<td>14.1</td>
</tr>
<tr>
<td>Model 6</td>
<td>$i^2i^3i^4i^5$</td>
<td>26.1</td>
<td>20.2</td>
<td>13.1</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.3</td>
<td>21.8</td>
<td>13.3</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.5</td>
<td>21.5</td>
<td>13.7</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.5</td>
<td>20.4</td>
<td>13.4</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.0</td>
<td>21.6</td>
<td>12.8</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23.9</td>
<td>20.3</td>
<td>12.5</td>
<td>8.7</td>
</tr>
</tbody>
</table>