Lecture 25: Syntactic Translation

Dan Klein – UC Berkeley

Stats from Liang Huang and Jonathan May

**MT Overview**

- **Phrase-Based MT**
  - synchronous context-free grammars (SCFGs)
  - context-free grammar in two dimensions
  - generating pairs of strings/trees simultaneously
  - co-indexed nonterminal further rewritten as a unit

- **Syntax-Based MT**
  - translation with SCFGs => monolingual parsing
  - parse the source input with the source projection
  - build the corresponding target sub-strings in parallel
Compact Forests

Language Model Costs

<table>
<thead>
<tr>
<th></th>
<th>1.0</th>
<th>3.0</th>
<th>8.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>(VP \text{hold} \ast \text{meeting})</td>
<td>1.0</td>
<td>2.0</td>
<td>4.0 + 5.0</td>
</tr>
<tr>
<td>(VP \text{hold} \ast \text{talk})</td>
<td>1.1</td>
<td>2.1 + 0.5</td>
<td>4.1 + 5.4 + 9.1 + 0.3</td>
</tr>
<tr>
<td>(VP \text{hold} \ast \text{conference})</td>
<td>3.5</td>
<td>4.5 + 0.6 + 6.5 + 10.5 + 11.5 + 0.6</td>
<td></td>
</tr>
</tbody>
</table>

k-best parsing
(Huang and Chang, 2005)

\begin{itemize}
\item a priority queue of candidates
\item extract the best candidate
\item push the two successors
\end{itemize}
### Language Model Costs

[Huang and Chiang 06]

Items are popped out-of-order.

**Solution:** keep a buffer of pop-ups

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5</td>
<td>2.4</td>
<td>5.1</td>
</tr>
<tr>
<td>(VP hold + meeting)</td>
<td>1.0</td>
<td>2.5</td>
<td>9.0</td>
</tr>
<tr>
<td>(VP hold + talk)</td>
<td>1.1</td>
<td>2.4</td>
<td>9.5</td>
</tr>
<tr>
<td>(VP hold + conference)</td>
<td>3.5</td>
<td>5.1</td>
<td>17.0</td>
</tr>
</tbody>
</table>

### Learning MT Grammars

- **Syntax-directed, English to Chinese** (Huang, Knight, Joshi, 2006)
- **First parse input, and then recursively transfer**

#### Synchronous tree-substitution grammars (STSG)

(Galley et al., 2004; Eisner, 2003)

### Idealistic machine translation

- Theorize a translation model
- Train on parallel sentences with unsupervised learning methods
- Examples: Berger et. al ’94 (IBM Model 4), Wu ’96 (ITG), Yamada and Knight ’01

### Realistic machine translation

- A relatively simple model is used to learn alignments unsupervisedly
- The alignments are used to limit the exploration space of a more complicated model

### Realistic machine translation

- The complicated model is trained on a smaller set of sentence pairs
- Counting and smoothing is fairly quick

### Realistic machine translation

- Empirically shown to be better than idealistic approach
- Scales well
- Examples: Och & Ney ’04, Chiang ’05, Galley et. al ’06
Realistic machine translation

- New research is on advanced models, but older models are used as bootstrap
- Can we add the power in our supervised model to the unsupervised component?

Extracting syntactic rules

- Obtain alignments
- Extract rules (Galley et al. '04, '06)
- Rules can...
  - capture phrasal translation
  - reorder parts of the tree
  - traverse the tree without reordering
  - insert (and delete) words

Bad alignments make bad rules

This isn't very good, but let's look at a worse example...

Sometimes they're really bad

One bad link makes a totally unusable rule!
**Rule Learning?**

- Rule composition
- Tree transforms (binarization, etc)
- Crazy pruning

**Other Important Issues**

**Results**

From [DeNeefe et al 07]

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Clause Dev</th>
<th>Test Dev</th>
<th>Arc tic Dev</th>
<th>Test Arc tic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline ATS</td>
<td>34.94</td>
<td>32.83</td>
<td>50.46</td>
<td>50.52</td>
</tr>
<tr>
<td>Baseline GJKM (minimal only)</td>
<td>38.02</td>
<td>37.67</td>
<td>49.34</td>
<td>49.99</td>
</tr>
<tr>
<td>GJKM composed size 2</td>
<td>40.24</td>
<td>39.75</td>
<td>50.76</td>
<td>50.94</td>
</tr>
<tr>
<td>GJKM composed size 3</td>
<td>40.95</td>
<td>40.44</td>
<td>51.56</td>
<td>51.48</td>
</tr>
<tr>
<td>GJKM composed size 4</td>
<td>41.36</td>
<td>40.69</td>
<td>51.60</td>
<td>51.71</td>
</tr>
<tr>
<td>GJKM minimal + SPMT model 1</td>
<td>39.78</td>
<td>39.16</td>
<td>50.17</td>
<td>51.27</td>
</tr>
<tr>
<td>GJKM composed + SPMT model 1</td>
<td>42.04</td>
<td>41.07</td>
<td>51.73</td>
<td>51.53</td>
</tr>
<tr>
<td>With beam search</td>
<td>42.17</td>
<td>41.26</td>
<td>52.56</td>
<td>53.70</td>
</tr>
</tbody>
</table>