Statistical Syntax-Directed Translation

with Extended Domain of Locality

Liang Huang (Penn), Kevin Knight (ISI), Aravind Joshi (Penn)

AMTA 2006, Boston, MA
Syntax-Directed Translation

• How do we human-beings do translation?
  1. understand the source sentence
  2. generate the target sentence

• Compiling
  1. parse input program into an expression/syntax tree
  2. generate code in machine language

• Machine Translation?
  1. parse the source sentence into a tree
  2. recursively transfer the tree into the target language

  • this work
Recursive Transfer

- converting tree-fragments into the target language

using the rules of (Galley et al., 2004)
Recursive Transfer

- converting tree-fragments into the target language
Recursive Transfer

• converting tree-fragments into the target language

槍手

_qiangshou_

Gunman
Recursive Transfer

- converting tree-fragments into the target language

枪手 被

qiangshou bei

gunman passive

NP-C

<table>
<thead>
<tr>
<th>DT</th>
<th>NN</th>
<th>VBN</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>police</td>
<td>shot</td>
</tr>
</tbody>
</table>

VP

<table>
<thead>
<tr>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TO</td>
</tr>
<tr>
<td>NP-C</td>
</tr>
</tbody>
</table>

| death |

<table>
<thead>
<tr>
<th>NP-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
</tr>
<tr>
<td>to</td>
</tr>
</tbody>
</table>
Recursive Transfer

- converting tree-fragments into the target language

### Formal Framework

- I-state
- extended left-hand-side
- linear and non-deleting
- root-to-frontier tree transducer
- \( (1-xRLN) \)

### Examples

枪手 被 警察 击毙。

- qiangshou bei jingcha jibi
- gunman passive police kill

(Graehl and Knight, 2004)
The gunman was killed by police.

Comparison with ISI system

ISI: decoding

ours: encoding
Comparison with SCFGs

(Chiang, 2005):

(was X₁ by X₂, 被 X₂ X₁)

I was here by myself

我 被 我 自己 在 这里

I was asleep by 10 pm

我 被 10 点 睡着
Search

- depth-first-search (DFS)
  - for each tree node
  - try all rules applicable
  - recursion on subtrees
  - plugin the results
- many overlapping rules
  - exponential run-time!
- solution: memoization
  - every node visited once
  - linear-time algorithm
Adding Language Models

- Generate top-$k$ translations
  - then rescore with language model
- fast, using Algorithm 3 of (Huang and Chiang, 2005)
- but many duplicate translations
  - due to spurious ambiguity
- one solution: determinization (May and Knight, 2006)
Non-Duplicate $k$-best

- a simple modification to the lazy $k$-best algorithm
- at each node, store a list of “unique solutions”
- keep asking for next-best until you get something new
- duplicates eliminated recursively at sub-problems
- here “uniqueness” can be any equivalence relation

**duplicate ratio = 1:40**

with our trick:
getting 100k unique is only about twice as slow as 100k non-unique
Experiments: English-to-Chinese

- **Training data**
  - Chinese/English parallel data from newswire
  - 1.95 M sentence pairs, English side: 28.3 M tokens
  - 24.7 M tree-to-string rules extracted (Galley et al., 2004)
  - Chinese trigram model trained on the Chinese side

- **Evaluation data**
  - NIST 2003 evaluation set (Chinese-to-English)
  - subset: 140 short sentences (≤ 25 Chinese words)
  - pick the first and second English references as source
  - equally divide into dev and test (each w/ 140 sent.)
  - single Chinese reference for each input sentence
Systems

- **Baseline:** Pharaoh (Koehn, 2004)
  - log-linear model of eight features
  - feature weights tuned by max-bleu algorithm (Och, 2003)

- **This work**
  - build the TM derivation forest (linear-time algorithm)
  - get top-$k$ unique translations
  - rescore with the trigram model
  - log-linear model of 3 features: TM, LM, length penalty
  - fix $w_{TM}$, grid-search $w_{LM}$, and binary-search len-penalty
## Results

- character-based BLEU (Chinese output), 95% interval

<table>
<thead>
<tr>
<th>System</th>
<th>dev set BLEU-4</th>
<th>test set (140 sentences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharaoh (with max-BLEU tuning)</td>
<td>25.96 ±2.8</td>
<td>23.54 ±1.9</td>
</tr>
<tr>
<td>direct model (1-best)</td>
<td>22.10 ±2.6</td>
<td>24.53 ±2.2</td>
</tr>
<tr>
<td>log-linear model (rescoring non-duplicate $k$-best list)</td>
<td>26.01 ±2.7</td>
<td>25.74 ±2.3</td>
</tr>
<tr>
<td>$k = 5000$</td>
<td>26.95 ±2.8</td>
<td>26.69 ±2.4</td>
</tr>
<tr>
<td>$k = 50000$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\beta = 0.994, \lambda = 0.513$

$p<0.05 \quad p<0.01$

sign-test (Collins et al., 2005) over Pharaoh

both BLEU-4 and BLEU-8 significantly better than Pharaoh on test-set
<table>
<thead>
<tr>
<th>source</th>
<th>the small town of wertheim is expected to experience the highest water level in 80 years.</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference</td>
<td>小镇 韦尔特海姆 预计 将 经历 80 年 来 的 最高 水位。</td>
</tr>
<tr>
<td>pharaoh</td>
<td>小 城镇, 预计 最高 水位 。 80 年 的 经验 韦 特 海 姆</td>
</tr>
<tr>
<td>l-best TM</td>
<td>韦尔特海姆 小镇 预计 经历 最高 水位 在 80 年。</td>
</tr>
<tr>
<td>rescored</td>
<td>韦尔特海姆 小镇 将 经历 80 年 的 最高 水位。</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>source</th>
<th>the global economic rebound and sustained economic growth in china last year have infused fresh vitality into the development of the port of shanghai.</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference</td>
<td>去年 全球 经济 回 暖， 中国 经济 持续 发展， 给 上海 港 的 发展 注入 新 的 活力。</td>
</tr>
<tr>
<td>pharaoh</td>
<td>世界 经济 的 回升 , 经济 持续 增长 注入 新 的 活力 中国 去年 的 港口 发展 到 上海 。</td>
</tr>
<tr>
<td>l-best TM</td>
<td>去年 中国 全球 经济 回升 与 经济 持续 增长 在 有 注入 新 的 活力 为 上海 港口 的 发展 。</td>
</tr>
<tr>
<td>rescored</td>
<td>去年 中国 全球 经济 回升 和 经济 持续 增长 为 上海 港口 的 发展 注入 新 的 活力。</td>
</tr>
</tbody>
</table>