Forest-to-String Statistical Translation Rules

Yang Liu, Qun Liu, and Shouxun Lin
Institute of Computing Technology
Chinese Academy of Sciences
Outline

- Introduction
- Forest-to-String Translation Rules
- Training
- Decoding
- Experiments
- Conclusion
Syntactic and Non-syntactic Bilingual Phrases

President Bush made a speech

non-syntactic

syntactic
Importance of Non-syntactic Bilingual Phrases

- About 28% of bilingual phrases are non-syntactic on an English-Chinese corpus (Marcu et al., 2006).
- Requiring bilingual phrases to be syntactically motivated will lose a good amount of valuable knowledge (Koehn et al., 2003).
- Keeping the strengths of phrases while incorporating syntax into statistical translation results in significant improvements (Chiang, 2005).
Previous Work

Galley et al., 2004
Previous Work

Marcu et al., 2006

*NPB_*NN

DT JJ NN

the mutual understanding

THE MUTUAL UNDERSTANDING
Previous Work

Liu et al., 2006

NP
  NP  VP
  NR  NN  VV  NN
  BUSH  PRESIDENT  MADE  SPEECH

NP
  NR  NN
  BUSH  PRESIDENT

President  Bush
Our Work

- We augment the tree-to-string translation model with
  - *forest-to-string rules* that capture non-syntactic phrase pairs
  - *auxiliary rules* that help integrate forest-to-string rules into the tree-to-string model
Outline

- Introduction
- Forest-to-String Translation Rules
- Training
- Decoding
- Experiments
- Conclusion
Tree-to-String Rules
A derivation is a left-most composition of translation rules that explains how a source parse tree, a target sentence, and the word alignment between them are synchronously generated.
A Derivation Composed of TRs
A Derivation Composed of TRs
A Derivation Composed of TRs
A Derivation Composed of TRs

- IP
  - NP
    - NN: GUNMAN
  - VP
    - SB
    - VP
      - WAS
      - NP
        - VV
      - NN: KILLED
    - NN: KILLED

- VP
  - SB
  - VP
    - WAS
    - NP
      - NN: KILLED
    - NN: KILLED

- the gun man was killed by [Diagram Insert]
- was killed by [Diagram Insert]
A Derivation Composed of TRs

the gunman was killed by police
A Derivation Composed of TRs

the gunman was killed by police.
Forest-to-String and Auxiliary Rules

forest = tree sequence!

care about only root sequence while incorporating forest rules
A Derivation Composed of TRs, FRs, and ARs
A Derivation Composed of TRs, FRs, and ARs
A Derivation Composed of TRs, FRs, and ARs
A Derivation Composed of TRs, FRs, and ARs

the gunman was killed by

the gunman was killed by

killed by
A Derivation Composed of TRs, FRs, and ARs

INSTITUTE OF COMPUTING TECHNOLOGY

A Derivation Composed of TRs, FRs, and ARs

INSTITUTE OF COMPUTING TECHNOLOGY

A Derivation Composed of TRs, FRs, and ARs
Outline

- Introduction
- Forest-to-String Translation Rules
- Training
- Decoding
- Experiments
- Conclusion
Training

- Extract both *tree-to-string* and *forest-to-string* rules from word-aligned, source-side parsed bilingual corpus
- Bottom-up strategy
- *Auxiliary* rules are NOT learnt from real-world data
An Example

President Bush made a speech.

Tree representation:

```
NP
  NP
    NP
      NR  NN  VV  NN
        BUSH  PRESIDENT  MADE  SPEECH
    VP
      NR
        BUSH

President Bush made a speech
```
An Example

President Bush made a speech.
An Example

President Bush made a speech.
An Example

President Bush made a speech.
An Example

INSTITUTE OF COMPUTING TECHNOLOGY

President Bush made a speech.

President

Bush

BUSH

PRESIDENT

MADE

SPEECH

President

Bush

BUSH

PRESIDENT

made a speech

President

Bush

BUSH

PRESIDENT

President

Bush
President Bush made a speech.
An Example

President Bush made a speech.

Diagram: A tree structure illustrating the parse of the sentence "President Bush made a speech."
An Example

President Bush made a speech.

10 FRs
An Example

President Bush made a speech
An Example

President Bush made a speech

max_height = 2
The development of Shanghai’s Pudong is in step with the establishment of its legal system.
Outline

- Introduction
- Forest-to-String Translation Rules
- Training
- Decoding
- Experiments
- Conclusion
Decoding

- Input: a source parse tree
- Output: a target sentence
- Bottom-up strategy
- Build auxiliary rules while decoding
- Compute subcell divisions for building auxiliary rules
An Example

Rule
- NR
  - BUSH
  - Bush

Derivation
- ( NR BUSH ) ||| Bush ||| 1:1

Translation
- Bush
An Example

Rule

NN
PRESIDENT
President

Derivation

( NN PRESIDENT ) ||| President ||| 1:1

Translation

President
An Example

Rule

VV
MADE
made

Derivation

( VV MADE ) ||| made ||| 1:1

Translation

made
An Example

Rule

NN

SPEECH

Derivation

( NN SPEECH ) ||| speech ||| 1:1

Translation

speech
An Example

```
(BUSH PRESIDENT) ||| Bush ||| 1:1
(NN PRESIDENT) ||| President ||| 1:1
```

**Rule**

```
NP
   /    \
  /      \\   
NR    NN

President Bush
```
An Example

Rule

Derivation

Translation
An Example

NP

NP

VP

NP

NN

VV

NN

BUSH

PRESIDENT

MADE

SPEECH

Rule

VP

VV

NN

MADE

made a

Derivation

( VP ( VV MADE ) ( NN ) ) ||| made a X ||| 1:1 2:3
( NN SPEECH ) ||| speech ||| 1:1

Translation

made a speech
An Example

Institute of Computing Technology

Rule

Translation

President Bush made a
An Example

Rule

Derivation

Translation
An Example

Rule

Derivation

Translation

President Bush made a speech
Subcell Division

NP

NP

NP

NN

VR

NN

BUSH

PRESIDENT

MADE

SPEECH

1:1 2:2 3:3 4:4
Subcell Division

NP
  NP  VP
    NR  NN  VV  NN
      BUSH  PRESIDENT  MADE  SPEECH

1:3 4:4
Subcell Division

NP
  /   |
 NP   VP
  /    |
 NR   NN
  |
 BUSH PRESIDENT MADE SPEECH

1:4
1:1 2:4
1:2 3:4
1:3 4:4
1:1 2:2 3:4
1:1 2:3 4:4
1:2 3:3 4:4
1:1 2:2 3:3 4:4

2^{(n-1)}
Build Auxiliary Rule

NP
  NP
  NR  NN
  BUSH  PRESIDENT

VP
  VV  NN
  MADE  SPEECH

NP
  NP
  NR  NN  VV  NN
  BUSH  PRESIDENT  MADE  SPEECH
Penalize the Use of FRs and ARs

- Auxiliary rules, which are built rather than learnt, have no probabilities.
- We introduce a feature that sums up the node count of auxiliary rules to balance the preference between
  - conventional tree-to-string rules
  - new forest-to-string and auxiliary rules
Outline

- Introduction
- Forest-to-String Translation Rules
- Training
- Decoding
- Experiments
- Conclusion
Experiments

- Training corpus: 31,149 sentence pairs with 843K Chinese words and 949K English words
- Development set: 2002 NIST Chinese-to-English test set (571 of 878 sentences)
- Test set: 2005 NIST Chinese-to-English test set (1,082 sentences)
Tools

- Evaluation: mteval-v11b.pl
- Language model: SRI Language Modeling Toolkits (Stolcke, 2002)
- Significant test: Zhang et al., 2004
- Parser: Xiong et al., 2005
- Minimum error rate training: optimizeV5IBMBLEU.m (Venugopal and Vogel, 2005)
## Rules Used in Experiments

<table>
<thead>
<tr>
<th>Rule</th>
<th>L</th>
<th>P</th>
<th>U</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>251,173</td>
<td>0</td>
<td>0</td>
<td>251,173</td>
</tr>
<tr>
<td>TR</td>
<td>56,983</td>
<td>41,027</td>
<td>3,529</td>
<td>101,539</td>
</tr>
<tr>
<td>FR</td>
<td>16,609</td>
<td>254,346</td>
<td>25,051</td>
<td>296,006</td>
</tr>
</tbody>
</table>
# Comparison

<table>
<thead>
<tr>
<th>System</th>
<th>Rule Set</th>
<th>BLEU4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharaoh</td>
<td>BP</td>
<td>0.2182 ± 0.0089</td>
</tr>
<tr>
<td>Lynx</td>
<td>BP</td>
<td>0.2059 ± 0.0083</td>
</tr>
<tr>
<td></td>
<td>TR</td>
<td>0.2302 ± 0.0089</td>
</tr>
<tr>
<td></td>
<td>TR + BP</td>
<td>0.2346 ± 0.0088</td>
</tr>
<tr>
<td></td>
<td>TR + FR + AR</td>
<td>0.2402 ± 0.0087</td>
</tr>
</tbody>
</table>
TRs Are Still Dominant

To achieve the best result of 0.2402, Lynx made use of:
- 26,082 tree-to-string rules
- 9,219 default rules
- 5,432 forest-to-string rules
- 2,919 auxiliary rules
## Effect of Lexicalization

<table>
<thead>
<tr>
<th>Forest-to-String Rule Set</th>
<th>BLEU4</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.2225 ± 0.0085</td>
</tr>
<tr>
<td>L</td>
<td>0.2297 ± 0.0081</td>
</tr>
<tr>
<td>P</td>
<td>0.2279 ± 0.0083</td>
</tr>
<tr>
<td>U</td>
<td>0.2270 ± 0.0087</td>
</tr>
<tr>
<td>L + P + U</td>
<td>0.2312 ± 0.0082</td>
</tr>
</tbody>
</table>
Outline

I Introduction
I Forest-to-String Translation Rules
I Training
I Decoding
I Experiments
I Conclusion
Conclusion

- We augment the tree-to-string translation model with
  - forest-to-string rules that capture non-syntactic phrase pairs
  - auxiliary rules that help integrate forest-to-string rules into the tree-to-string model
- Forest and auxiliary rules enable tree-to-string models to derive in a more general way and bring significant improvement.
Future Work

- Scale up to large data
- Further investigation in auxiliary rules
Thanks!