New Advances in Forest-based Translation

Haitao Mi

Institute of Computing Technology
Chinese Academy of Sciences
Tree-to-string Translation

Input: 布什与沙龙举行会谈

(Liu et al., 2006; Huang et al., 2006)
Tree-to-string Translation

Segmentation

Input

布什 与 沙龙 举行 会谈

布什与沙龙举行会谈

(Liu et al., 2006; Huang et al., 2006)
Tree-to-string Translation

Parsing

Segmentation

Input

布什与沙龙举行会谈

(Liu et al., 2006; Huang et al., 2006)
Tree-to-string Translation

Translation

Parsing

Segmentation

Input

布什与沙龙举行会谈

(Liu et al., 2006; Huang et al., 2006)
Bush hold a talk with Sharon

(Liu et al., 2006; Huang et al., 2006)
Tree-to-string Translation

Translation

Parsing

Segmentation

Input  布什与沙龙举行会谈

(Liu et al., 2006; Huang et al., 2006)
Bush and Sharon hold a talk

布什与沙龙举行会谈

(Liu et al., 2006; Huang et al., 2006)
Forest-to-string Translation

Translation

Parsing

Segmentation

Input

(Mi, Huang, and Liu. 2008; Mi and Huang. 2008)
Forest-to-string Translation

Translation

Parsing

Segmentation

Input

(Mi, Huang, and Liu. 2008; Mi and Huang. 2008)
Forest-to-string Translation

Translation

Bush hold a talk with Sharon

Parsing

Bush and Sharon hold a talk

Segmentation

布什与沙龙举行会谈

Input

布什与沙龙举行会谈

(Mi, Huang, and Liu. 2008; Mi and Huang. 2008)
New Advances
New Advances

- Larger Search Space
- More Powerful Grammar
- Faster Search Algorithm
- Context Sensitive Model
- Dependency Trees
New Advances

- Larger Search Space
  - Translation with Lattices and Forests (Mi, Huang, and Liu. 2010)
- More Powerful Grammar
- Faster Search Algorithm
- Context Sensitive Model
- Dependency Trees
Translation with **Lattices** and **Forests**

Translation: Bush hold a talk with Sharon

Parsing

Segmentation

Input: 布什与沙龙举行会谈

(Mi, Huang, and Liu. 2010) COLING 2010
Translation with **Lattices** and Forests

Translation

Bush hold a talk with Sharon

Parsing

Lattice

Input

布什与沙龙举行会谈

(Mi, Huang, and Liu. 2010) COLING 2010
Translation with **Lattices** and Forests

Translation

Bush hold a talk with Sharon

Parsing

Lattice

Input

布什与沙龙举行会谈

(Mi, Huang, and Liu. 2010) COLING 2010
Translation with Lattices and Forests

Translation

Bush hold a talk with Sharon

Parsing

Lattice

Input

布什与沙龙举行会谈

(Mi, Huang, and Liu. 2010) COLING 2010
Experiments
Data

- Chinese-to-English translation
- FBIS data (7M/9M Chinese/English words)
  - Chinese
    - Seg&POS tagging system (Jiang et al., 08)
    - Parser (Xiong et al., 05)
    - Marginal probability-based pruning algorithm (Huang 08)
- GIZA++: <most-refined segmentation, string>
- 4-gram LM trained on 1/3 portion of Giga Xinhua
- dev: NIST 2002 (878 sent.)
- test: NIST 2005 (1082 sent.)

(Mi, Huang, and Liu. 2010) COLING 2010
## Results

<table>
<thead>
<tr>
<th>system</th>
<th># of rules</th>
<th></th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
<td>dev &amp; tst</td>
<td></td>
</tr>
<tr>
<td>forest</td>
<td>29.6M</td>
<td>3.3M</td>
<td>28.75</td>
</tr>
<tr>
<td>lattice &amp; forest</td>
<td>23.5M</td>
<td>3.4M</td>
<td>29.65*</td>
</tr>
</tbody>
</table>

(Mi, Huang, and Liu. 2010) COLING 2010
New Advances

• Larger Search Space
  • Translation with Lattices and Forests (Mi, Huang, and Liu. 2010)

• More Powerful Grammar
  • Constituency-to-Dependency Grammar (Mi and Liu. 2011)

• Faster Search Algorithm

• Context Sensitive Model

• Dependency Trees
Syntax-based SMT Systems
Syntax-based SMT Systems

String-to-Tree   Tree-to-String   Tree-to-Tree
<table>
<thead>
<tr>
<th>string-to-constituency</th>
<th>constituency-to-string</th>
<th>constituency-to-constituency</th>
</tr>
</thead>
<tbody>
<tr>
<td>string-to-dependency</td>
<td>dependency-to-string</td>
<td>dependency-to-dependency</td>
</tr>
<tr>
<td>String-to-constituency</td>
<td>Constituency-to-string</td>
<td>Constituency-to-constituency</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>String-to-dependency</td>
<td>Dependency-to-string</td>
<td>Dependency-to-dependency</td>
</tr>
</tbody>
</table>

Fast Grammaticality

BLEU
Fast Grammaticality BLEU

string-to-constituency  constituency-to-string  constituency-to-constituency

string-to-dependency  dependency-to-string  dependency-to-dependency
Bush held a talk with Sharon.

**Fast Grammaticality**

- **String-to-constituency:** ❌
- **Constituency-to-string:** ✓
- **Constituency-to-constituency:** ✓

**BLEU**

- **String-to-dependency:** ✓
- **Dependency-to-string:** ✓
- **Dependency-to-dependency:** ✓
Bush held a talk with Sharon

Fast Grammaticality

Grammaticality

BLEU

String-to-constituency

Constituency-to-string

Constituency-to-constituency

String-to-dependency

Dependency-to-string

Dependency-to-dependency
string-to-constituency

constituency-to-string

constituency-to-constituency

Fast
Grammaticality
BLEU

string-to-dependency
dependency-to-string
dependency-to-dependency

bushi yu shalong juxing le huitan
Bush held a talk with Sharon

bushi yu shalong juxing le huitan
Bush held a talk with Sharon

bushi yu shalong juxing le huitan
Bush held a talk with Sharon
Fast Grammaticality

BLEU
Fast Grammaticality

string-to-constituency: ×

constituency-to-string: √

constituency-to-constituency: √

string-to-dependency: ×

dependency-to-string: √

dependency-to-dependency: ×
Constituency-to-Dependency

Fast ✓
Grammaticality ✓
BLEU ✓
Constituency-to-Dependency

(Mi and Liu. 2010) ACL 2010
Generative Story
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Generative Story

(Mi and Liu. 2010) ACL 2010
Bush held with Sharon a talk (Mi and Liu. 2010) ACL 2010
Bush held a talk with Sharon (Mi and Liu. 2010) ACL 2010
Bush held a talk with Sharon.
Rule Extraction

- Frontier set computation

(Mi and Liu. 2010) ACL 2010
Frontier set computation
• Frontier set computation

(Mi and Liu. 2010) ACL 2010
Bush held a talk with Sharon.
Rule Extraction

- Frontier set computation

(Mi and Liu. 2010) ACL 2010
Bush held a talk with Sharon.

(Mi and Liu. 2010) ACL 2010
Frontier set computation
Rule Extraction

- Fragmentation

(Mi and Liu. 2010) ACL 2010
Rule Extraction

- **Fragmentation**

(Mi and Liu. 2010) ACL 2010
Rule Extraction

- Fragmentation

Bush held a talk with Sharon

(Mi and Liu. 2010) ACL 2010
Bush held a talk with Sharon.

(Mi and Liu. 2010) ACL 2010
Bush held a talk with Sharon

(Mi and Liu. 2010) ACL 2010
Decoding

- Pattern-matching

Rules

Parse Forest - LM Translation Forest

(Mi and Liu. 2010) ACL 2010
Decoding

- Pattern-matching

Rules

Parse Forest - LM Translation Forest

(Mi and Liu. 2010) ACL 2010
Decoding

- Pattern-matching

Rules

NPB  
| bushi
↓
Bush

CC  
| yu
↓
with

Parse Forest

(Mi and Liu. 2010) ACL 2010

- LM Translation Forest
Decoding

- Pattern-matching

Rules

- Parse Forest
- LM Translation Forest

(Mi and Liu. 2010) ACL 2010
Decoding

- Pattern-matching

Rules

(Bush with Sharon talk)

(Mi and Liu. 2010) ACL 2010
Decoding

**Pattern-matching**

### Rules

- NPB: bushi
- CC: yu
- NPB: shalong
- NPB: huitan

### Parse Forest

- NPB0,1
  - CC1,2
  - P1,2
- NPB2,3
  - VS3,4
  - AS4,5
  - NPB5,6
- NPB0,3
  - PP1,3
  - VP1,6
- IP0,6

### LM Translation Forest

- VPB3,6
  - VPB3,6
  - NPB5,6

---

(Mi and Liu. 2010) ACL 2010
Decoding

- Pattern-matching

Rules

<table>
<thead>
<tr>
<th>NPB</th>
<th>bushi</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>yu</td>
</tr>
<tr>
<td>NPB</td>
<td>shalong</td>
</tr>
<tr>
<td>NPB</td>
<td>huitan</td>
</tr>
</tbody>
</table>

Parse Forest

(Mi and Liu. 2010) ACL 2010
Decoding

- +LM Decoding

Rules

Parse Forest

+LM Translation Forest

(Mi and Liu. 2010) ACL 2010

Bush with Sharon held ((a) talk) (with Sharon)

VPB hold ((a) talk)
Experiments
Data

- Task: Chinese-English
- FBIS data
  - (7M/9M Chinese/English words)
- Parsing
  - source side: parse forests (p=3)
  - target side: 1-best trees
- GIZA++ aligned
- LM
  - 4-gram surface string, trained on 1/3 portion of Giga Xinhua
  - 3-gram dependency, trained on English side of training data
- dev: NIST 2002 (878 sent.) (parser forest: p=10)
- test: NIST 2005 (1082 sent.) (parser forest: p=10)

(Mi and Liu. 2010) ACL 2010
## Results

<table>
<thead>
<tr>
<th>System</th>
<th>Rule</th>
<th>Type</th>
<th>#</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forest</strong></td>
<td></td>
<td>c2s+s2s</td>
<td>31.9M+77.9M</td>
<td>34.17</td>
</tr>
<tr>
<td>c2s</td>
<td></td>
<td>c2d+s2s</td>
<td>13.8M+ 9.0M</td>
<td>34.03(-0.1)</td>
</tr>
<tr>
<td>c2d+s2d</td>
<td></td>
<td>13.8M+77.9M</td>
<td>32.28(-1.7)</td>
<td></td>
</tr>
<tr>
<td>c2d+s2d-dep</td>
<td></td>
<td>13.8M+77.9M</td>
<td>33.25(-0.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Forest</strong></td>
<td></td>
<td>c2d+s2d</td>
<td>13.8M+ 9.0M</td>
<td>34.88(+0.7)</td>
</tr>
<tr>
<td>c2d</td>
<td></td>
<td>13.8M+77.9M</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Mi and Liu. 2010) ACL 2010
New Advances

- Larger Search Space
  - Translation with Lattices and Forests (Mi, Huang, and Liu. 2010)
- More Powerful Grammar
  - Constituency-to-Dependency Grammar (Mi and Liu. 2011)
- Faster Search Algorithm
  - Incremental Decoding (Huang and Mi, 2011)
- Context Sensitive Model
- Dependency Trees
## MT: Phrase-based vs. Syntax-based

<table>
<thead>
<tr>
<th></th>
<th>phrase</th>
<th>syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypotheses generation</td>
<td>left-to-right</td>
<td>bottom-up</td>
</tr>
<tr>
<td>in theory</td>
<td>exponential</td>
<td>polynomial</td>
</tr>
<tr>
<td>in practice</td>
<td>fast (linear with beam search)</td>
<td>slow (with LM)</td>
</tr>
</tbody>
</table>

(Huang and Mi, 2010) EMNLP 2010
### MT: Phrase-based vs. Syntax-based

<table>
<thead>
<tr>
<th></th>
<th>phrase</th>
<th>syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypotheses generation</td>
<td>left-to-right</td>
<td>bottom-up</td>
</tr>
<tr>
<td>in theory</td>
<td>exponential</td>
<td>polynomial</td>
</tr>
<tr>
<td>in practice</td>
<td>fast (linear with beam search)</td>
<td>slow (with LM)</td>
</tr>
</tbody>
</table>

Borrow phrase-based decoding for tree-to-string?

(Huang and Mi, 2010) EMNLP 2010
## MT: Phrase-based vs. Syntax-based

<table>
<thead>
<tr>
<th></th>
<th>phrase</th>
<th>syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypotheses generation</td>
<td>left-to-right</td>
<td>left-to-right</td>
</tr>
<tr>
<td>in theory</td>
<td>exponential</td>
<td>polynomial</td>
</tr>
<tr>
<td>in practice</td>
<td>fast</td>
<td>fast</td>
</tr>
<tr>
<td></td>
<td>(linear with beam search)</td>
<td>(with LM)</td>
</tr>
</tbody>
</table>

Borrow phrase-based decoding for tree-to-string?

(Huang and Mi, 2010) EMNLP 2010
Incremental for Tree-to-String

- key intuition: adapt coverage-vector idea
- tree coverage-vector: which sub-trees translated

(Huang and Mi, 2010) EMNLP 2010
Incremental for Tree-to-String

- key intuition: adapt coverage-vector idea
- tree coverage-vector: which sub-trees translated

expand arbitrary

(Huang and Mi, 2010) EMNLP 2010
Incremental for Tree-to-String

- key intuition: adapt coverage-vector idea
- tree coverage-vector: which sub-trees translated

(Huang and Mi, 2010) EMNLP 2010
Incremental for Tree-to-String

- key intuition: adapt coverage-vector idea
- tree coverage-vector: which sub-trees translated

(Huang and Mi, 2010) EMNLP 2010
Tree Coverage Vector as Stack

IP
  /\  
NPB  VP
  /\  
    NR  
      Bùshí

IP
  /\  
PP  VPB
  /\  
P  NPB  VV  AS  NPB
  /\  /\  
yŭ NR jǔxíng le NN
  /\  
Shālóng huìtán
Tree Coverage Vector as Stack

- stack (*active* derivation history): \[\varepsilon \rightarrow \text{IP} \] \[[\text{IP} \rightarrow \text{NPB} \bullet \text{VP}]\]
Tree Coverage Vector as Stack

- **stack (active derivation history):** \([\varepsilon \to \text{IP}] [\text{IP} \to \text{NPB} \cdot \text{VP}]\)
- **expand point** (blue point at the top of stack)
Tree Coverage Vector as Stack

- **stack** (*active* derivation history): \[ \varepsilon \rightarrow \text{IP} \] \[ \text{IP} \rightarrow \text{NPB} \cdot \text{VP} \]
- **expand point** (blue point at the top of stack)
- **three colors for nodes:**

---

![Tree Coverage Vector as Stack Diagram]
Tree Coverage Vector as Stack

- **stack** (*active* derivation history): \([\varepsilon \rightarrow \text{IP}] \rightarrow \text{NPB} \rightarrow \text{VP}\)
- **expand point** (blue point at the top of stack)
- **three colors for nodes:**
  - white (uncovered), grey (partially covered), and black (covered)

“I have finished NPB subtree but not started with VP subtree”
Example Incremental Decoding

[ε→<s> •IP </s>]

<s>

(IP
  NPB
  NR
  Bushi)

(VP
  PP
  P
  yu)

(VPB
  VV
  juxing)

(AS
  le)

(NPB
  NN
  huitan)

stack hypothesis

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
\begin{align*}
[\varepsilon & \rightarrow <s> \quad \text{IP} \quad <s> ] \\
[\text{IP} & \rightarrow \text{NPB} \quad \text{VP} ]
\end{align*}
\]

<s>

\[
\text{IP}(x_1: \text{NPB} \ x_2: \text{VP}) \rightarrow x_1 \ x_2
\]

action: predict (push)

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow \text{<s> } \text{IP } \text{/s>}] \quad [\text{IP} \rightarrow \text{NPB } \text{VP}] \quad [\text{NPB} \rightarrow \text{Bush}]
\]

\text{<s>}

\text{action: predict (push)}

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[ \varepsilon \rightarrow <s> \text{IP } <\text{/s}> ] [\text{IP} \rightarrow \bullet \text{NPB VP}] [\text{NPB} \rightarrow \text{Bush} \bullet ] \]

<s> Bush

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow \langle s \rangle \bullet \text{IP} \langle /s \rangle] \ [\text{IP} \rightarrow \bullet \text{NPB} \ \text{VP}]
\]

\langle s \rangle \text{ Bush}

\text{action: complete (pop)}

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow <s> \bullet \text{IP} \langle/s\rangle] \ [\text{IP} \rightarrow \text{NPB} \bullet \text{VP}]
\]

\(<s>\) Bush

\(<s>\) IP \langle/s\rangle

\text{stack}

\text{hypothesis}

\text{action: complete (pop)}

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[ \varepsilon \rightarrow \text{IP} \] \[ \text{IP} \rightarrow \text{NPB} \rightarrow \text{VP} \] \[ \text{VP} \rightarrow \text{held NPB with NPB} \]

\( \text{stack} \)

\( \text{hypothesis} \)

\( \text{action: predict (push)} \)

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[ \varepsilon \rightarrow <s> \bullet \text{IP} <</s> > \] \[ \text{[IP} \rightarrow \text{NPB} \bullet \text{VP}] \] \[ \text{[VP} \rightarrow \text{held} \bullet \text{NPB with NPB]} \]

\(<s> \) Bush held

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow <s> \quad \text{IP} \quad <!--s--> ] \quad [\text{IP} \rightarrow \text{NPB} \cdot \text{VP}] \quad [\text{VP} \rightarrow \text{held} \cdot \text{NPB} \text{ with } \text{NPB}] \quad [\text{NPB} \rightarrow \text{talks }]
\]

\(<s>\) Bush held

\((\text{Huang and Mi, 2010})\) EMNLP 2010
Example Incremental Decoding

\[
\text{[\varepsilon \rightarrow <s> \ IP <</s>] [IP \rightarrow NPB \cdot VP] [VP \rightarrow held \cdot NPB \text{ with } NPB] [NPB \rightarrow talks \cdot ]}
\]

<s> Bush held talks

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow <s> \quad \text{IP} \quad <s>] \quad [\text{IP} \rightarrow \text{NPB} \cdot \text{VP}] \quad [\text{VP} \rightarrow \text{held} \quad \text{NPB} \cdot \text{with} \quad \text{NPB}]
\]

\(<s>\) Bush held talks

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow <s> \text{ IP } <s>] \ [\text{ IP } \rightarrow \text{ NPB } \cdot \text{ VP}] \ [\text{ VP } \rightarrow \text{ held } \text{ NPB } \text{ with} \cdot \text{ NPB}]
\]

<s> Bush held talks with

*(Huang and Mi, 2010) EMNLP 2010*
Example Incremental Decoding

\[
\begin{align*}
\varepsilon & \rightarrow <s> \text{ IP } <s> \text{ } \tag{1} \\
\text{IP} & \rightarrow \text{ NP} . \text{ VP} \tag{2} \\
\text{VP} & \rightarrow \text{ held } \text{ NP} \text{ with } \text{ NP} \tag{3} \\
\text{NP} & \rightarrow \text{ Sharon} \tag{4}
\end{align*}
\]

\(<s>\text{ } \text{Bush} \text{ held talks with } \text{}</s>\)

(Huang and Mi, 2010) EMNLP 2010

action: predict (push)
Example Incremental Decoding

\[\varepsilon \rightarrow <s> \quad \text{IP} \quad </s> \] \[\text{IP} \rightarrow \text{NPB} \cdot \text{VP} \] \[\text{VP} \rightarrow \text{held} \quad \text{NPB} \quad \text{with} \cdot \text{NPB}[\text{NPB} \rightarrow \text{Sharon}]\]

<s> Bush held talks with Sharon

action: scan

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[ \varepsilon \rightarrow <s> \ IP \ <s> \] [IP \rightarrow NPB \cdot VP] [VP \rightarrow \text{held} \ NPB \text{ with } NPB.] \\

\textless s\textgreater \ Bush \ held \ talks \ with \ Sharon

\textit{action: complete (pop)}

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow <s> \text{NPB} <\varepsilon>] [\text{IP} \rightarrow \text{NPB} \text{VP}].
\]

\(<s>\) Bush held talks with Sharon

\(<s>\) IP <\s>
Example Incremental Decoding

[ε→<s> IP. </s>]

<s> Bush held talks with Sharon

action: complete (pop)

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

\[
[\varepsilon \rightarrow <s> \text{ IP } </s> . ]
\]

\(<s> \text{ Bush held talks with Sharon } </s>\)

(Huang and Mi, 2010) EMNLP 2010
Example Incremental Decoding

$[\varepsilon \rightarrow <s> \text{IP} \ <</s> \cdot]$

$<s>$ Bush held talks with Sharon $<</s>$

(Huang and Mi, 2010) EMNLP 2010
Experiments
Experimental Setup

- Chinese-to-English translation
  - on a Python implementation of tree-to-string system
- 1.5M sentence pairs (38M/32M words in Chn/Eng)
- dev: NIST 2006 (616 sent); test: NIST 2008 (691 sent)
- Chinese-side parsed by Berkeley parser (Petrov & Klein, 07)
- rules extracted using GHKM algorithm (Galley et al, 04; 06)
- trigram language model trained on the English side
- feature weights tuned using MERT (Och, 03)
Comparison with Moses

- incremental tree-to-string is linear-time in practice
Tree-to-string Systems

![Graph showing average decoding time versus sentence length with two sets of data: cube pruning and incremental.]
Comparison with Cube Pruning

- incremental is slightly faster than cube pruning
- note they are very different (orthogonal) techniques
- we envision their combination will be even faster
Comparison with Cube Pruning

- incremental is slightly faster than cube pruning
- note they are very different (orthogonal) techniques
- we envision their combination will be even faster
Comparison with Cube Pruning

- incremental is slightly faster than cube pruning
- note they are very different (orthogonal) techniques
  - we envision their combination will be even faster
New Advances

- Larger Search Space
  - Translation with Lattices and Forests (Mi, Huang, and Liu. 2010)
- More Powerful Grammar
  - Constituency-to-Dependency Grammar (Mi and Liu. 2011)
- Faster Search Algorithm
  - Incremental Decoding (Huang and Mi, 2011)
- Context Sensitive Model
  - Rule Markov Model (Ashish, Mi, Huang, and Chiang. 2011)
- Dependency Trees
**Tree-to-string Translation**

<table>
<thead>
<tr>
<th>rules</th>
<th>#</th>
<th>decoding</th>
<th>BLEU</th>
</tr>
</thead>
</table>

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
## Tree-to-string Translation

<table>
<thead>
<tr>
<th>rules</th>
<th>#</th>
<th>decoding</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimal</td>
<td>small</td>
<td>fast</td>
<td>low</td>
</tr>
</tbody>
</table>

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
### Tree-to-string Translation

<table>
<thead>
<tr>
<th>rules</th>
<th>#</th>
<th>decoding</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimal</td>
<td>small</td>
<td>fast</td>
<td>low</td>
</tr>
<tr>
<td>composed</td>
<td>huge</td>
<td>slow</td>
<td>high</td>
</tr>
</tbody>
</table>
# Tree-to-string Translation

<table>
<thead>
<tr>
<th>rules</th>
<th>#</th>
<th>decoding</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimal</td>
<td>small</td>
<td>fast</td>
<td>low</td>
</tr>
<tr>
<td>composed</td>
<td>huge</td>
<td>slow</td>
<td>high</td>
</tr>
<tr>
<td>rule Markov Model</td>
<td>small</td>
<td>fast</td>
<td>high</td>
</tr>
</tbody>
</table>

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Generative Story
Generative Story

$IP^@\epsilon$

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Generative Story

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Generative Story

$P(r_2|r_1)$

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Generative Story

\[ P(r_3|r_1) \]

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Bush held talks

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Bush held talks
Bush held talks with NP@2.1.2

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Bush held talks with Sharon

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Probability of a Derivation Tree

\[
P(T) = P(r_1|\varepsilon) \cdot P(r_2|r_1) \cdot P(r_3|r_1) \\
P(r_5|r_1,r_3) \cdot P(r_4|r_1,r_3) \\
P(r_6|r_1,r_3,r_4) \cdot P(r_7|r_1,r_3,r_4)
\]

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Probability of a Derivation Tree

$$P(T) = P(r_1|\varepsilon) \cdot P(r_2|r_1) \cdot P(r_3|r_1)$$
$$P(r_5|r_1,r_3) \cdot P(r_4|r_1,r_3)$$
$$P(r_6|r_1,r_3,r_4) \cdot P(r_7|r_1,r_3,r_4)$$

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Probability of a Derivation Tree

\[ P(T) = P(r_1|\varepsilon) \cdot P(r_2|r_1) \cdot P(r_3|r_1) \cdot P(r_5|r_1,r_3) \cdot P(r_4|r_1,r_3) \cdot P(r_6|r_1,r_3,r_4) \cdot P(r_7|r_1,r_3,r_4) \]

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Probability of a Derivation Tree

\[ P(T) = P(r_1|\varepsilon) \cdot P(r_2|r_1) \cdot P(r_3|r_1) \cdot P(r_5|r_1, r_3) \cdot P(r_4|r_1, r_3) \cdot P(r_6|r_1, r_3, r_4) \cdot P(r_7|r_1, r_3, r_4) \]

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
\[
P(T) = P(r_1|\varepsilon) \cdot P(r_2|r_1) \cdot P(r_3|r_1) \\
P(r_5|r_1,r_3) \cdot P(r_4|r_1,r_3) \\
P(r_6|r_1,r_3,r_4) \cdot P(r_7|r_1,r_3,r_4)
\]

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Probability of a Derivation Tree

\[ P(T) = P(r_1|\varepsilon) \cdot P(r_2|r_1) \cdot P(r_3|r_1) \cdot P(r_5|r_1, r_3) \cdot P(r_4|r_1, r_3) \cdot P(r_6|r_1, r_3, r_4) \cdot P(r_7|r_1, r_3, r_4) \]

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Probability of a Derivation Tree

\[
P(T) = P(r_1|\varepsilon) \cdot P(r_2|r_1) \cdot P(r_3|r_1) \\
P(r_5|r_1,r_3) \cdot P(r_4|r_1,r_3) \\
P(r_6|r_1,r_3,r_4) \cdot P(r_7|r_1,r_3,r_4)
\]

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Learning Rule Markov Models

Galley et al.

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Learning Rule Markov Models

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Learning Rule Markov Models

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Learning Rule Markov Models

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
How to do it?

- composed rules
- combine minimal rules
How to do it?

- composed rules
- combine minimal rules
- use independently
How to do it?

- composed rules
  - combine minimal rules
  - use independently
- rule Markov model (RMM)
  - model the composition
  - generate rules conditioned on their ancestors
    - only take into account their ancestors (vertical context)
    - \[ P(T) = \prod_{r \in T} P(r \mid \text{anc}(r)) \]
How to do it?

- composed rules
- combine minimal rules
- use independently
- rule Markov model (RMM)
- model the composition
- generate rules conditioned on their ancestors
  - only take into account their ancestors (vertical context)

\[
P(T) = \prod_{r \in T} P(r \mid \text{anc}(r))
\]

\[
= P(r_1 \mid \varepsilon) \cdot P(r_2 \mid r_1) \cdot P(r_3 \mid r_1) \cdot P(r_4 \mid r_1) \cdot P(r_5 \mid r_1) \cdot P(r_6 \mid r_1, r_2)
\]
Incremental Decoding with RMM

very easy!

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

![Diagram of derivation tree]

derivation tree $r_1$

$P(r_1|\epsilon)$

action: predict (push)

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

\[ P(r_2|r_1) \]

Derivation tree: \( r_1 \) \( \xrightarrow{r_2} \)

\(<s> \) IP \(</s>\)

```
<s> IP </s>
```

**Action:** predict (push)

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

Derivation tree:

\[ r_1 \]

\[ r_2 \]

Action: scan

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

action: complete (pop)

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

\[
P(r_3 | r_1)
\]

 derivation tree

\[
  r_1 \quad r_2 \quad r_3
\]

action: predict (push)

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(action: scan)

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

\[
P(r_4|r_1,r_3)
\]

\[
\text{derivation tree}
\]

Action: predict (push)

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Incremental Decoding with RMM

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Experiments
Data

- Task: Chinese to English
- Training: 1.5 million sentence pairs with 38/32 million words of Chinese/English
- Dev set: 616 sentences of the Newswire portion of 2006 NIST MT evaluation test set
- Test set: 619 sentences of the Newswire portion of 2006 NIST MT evaluation test set

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Results

- minimal
- minimal + trigram rule Markov model
- composed rule
- vertical composed rule

BLEU

- (Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Results

- context probabilities
- rules

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
## Results

![Graph showing context probabilities and rules]

<table>
<thead>
<tr>
<th>Minimal</th>
<th>Composed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500000</td>
<td></td>
</tr>
<tr>
<td>1000000</td>
<td></td>
</tr>
<tr>
<td>1500000</td>
<td></td>
</tr>
<tr>
<td>2000000</td>
<td></td>
</tr>
</tbody>
</table>

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Results

context probabilities

rules

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Results

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Results

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
Results

(Ashish, Mi, Huang, and Chiang. 2011) ACL 2011
New Advances

- Larger Search Space
  - Translation with Lattices and Forests (Mi, Huang, and Liu. 2010)
- More Powerful Grammar
  - Constituency-to-Dependency Grammar (Mi and Liu. 2011)
- Faster Search Algorithm
  - Incremental Decoding (Huang and Mi, 2011)
- Context Sensitive Model
  - Rule Markov Model (Ashish, Mi, Huang, and Chiang. 2011)
- Dependency Trees
  - Dependency-to-String Translation (Xie, Mi, and Liu. 2011)
Dependency-to-string Translation

- Dependency VS Constituency
Dependency-to-string Translation

- Dependency VS Constituency

Diagram:
- 举行
- 布什
- 与
- 沙龙
- 会谈
Dependency-to-string Translation

- Dependency VS Constituency

Best phrasal cohesion (Fox, 2002)
Dependency-to-string Translation

- Dependency VS Constituency

Best phrasal cohesion (Fox, 2002)
Encode semantic relations directly
Dependency-to-string Translation

- Conventional approaches

Bush held a talk with Sharon
Dependency-to-string Translation

- Conventional approaches

Bush held a talk with Sharon
Dependency-to-string Translation

- Conventional approaches

Bush held a talk with Sharon
Dependency-to-string Translation

- Conventional approaches

Bush held a talk with Sharon

arbitrary treelet
need insert operation
Dependency-to-string Translation

- Conventional approaches

Bush held a talk with Sharon

arbitrary treelet

need insert operation
• Conventional approaches

Bush held a talk with Sharon

arbitrary treelet
need insert operation

complicate
low performance
Only model head-dependencies

Bush held a talk with Sharon (Xie, Mi, and Liu. 2011) EMNLP 2011
Only model head-dependencies

Bush held a talk with Sharon

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Only model head-dependencies

Bush held a talk

(Xie, Mi, and Liu. 2011) EMNLP 2011
• Only model head-dependencies

Bush held a talk with Sharon

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Only model head-dependencies

Bush held a talk

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Only model head-dependencies

Bush held a talk with Sharon

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Only model head-dependencies
- Rule coverage, POS tags

Bush held a talk with Sharon

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Only model head-dependencies
- Rule coverage, POS tags
Dependency-to-string Translation

- Decoding

(Xie, Mi, and Liu. 2011) EMNLP 2011
Bush held a talk.

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Decoding

Bush held a talk 与 P 沙龙 NR

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Decoding

Bush held a talk with Sharon

(Xie, Mi, and Liu. 2011) EMNLP 2011
Dependency-to-string Translation

- Decoding

Bush held a talk with Sharon

(Xie, Mi, and Liu. 2011) EMNLP 2011
Experiments
Experimental Setup

- Baselines
  - Constituency-to-String model (Liu et al., 2006)
  - Replication of HPB (Chiang, 2005)

- C2E translation

- Training: 1.5M sentence pairs from LDC

- Dev Set: NIST 2002

- Test Sets: NIST 2004 /2005

- Stanford Parser

(Xie, Mi, and Liu. 2011) EMNLP 2011
BLEU

(Xie, Mi, and Liu. 2011) EMNLP 2011
New Advances

- Larger Search Space
  - Translation with Lattices and Forests (Mi, Huang, and Liu. 2010)
- More Powerful Grammar
  - Constituency-to-Dependency Grammar (Mi and Liu. 2011)
- Faster Search Algorithm
  - Incremental Decoding (Huang and Mi, 2011)
- Context Sensitive Model
  - Rule Markov Model (Ashish, Mi, Huang, and Chiang. 2011)
- Dependency Trees
  - Dependency-to-String Translation (Xie, Mi, and Liu. 2011)
References

• Haitao Mi, Liang Huang and Qun Liu. Forest-based Translation. In Proceedings of ACL 2008, Columbus, OH.


• Liang Huang and Haitao Mi. Efficient Incremental Decoding for Tree-to-String Translation. In Proceedings of EMNLP 2010, Boston, USA, October.


Thanks!

http://mtgroup.ict.ac.cn/~mihaitao