Joint Parsing and Alignment with Weakly Synchronized Grammars

David Burkett, John Blitzer, and Dan Klein
{dburkett,blitzer,klein}@cs.berkeley.edu
Statistical Machine Translation Rules
Statistical Machine Translation Rules

Input

Whereas in 1992

Xinhua News Agency

(End)

Beijing, February 13 -
Statistical Machine Translation Rules

**Input**

Whereas in 1992

Xinhua News Agency

（End）

Beijing, February 13 -

**Output**

ADVP  →  ADVP

PP  →  Whereas  PP

PP  →  in  CD

NT  →  1992
Beijing, February 13 -

Statistical Machine Translation Rules

Whereas in 1992

Xinhua News Agency

（End）

北京, 二月 十三日 电
Beijing, February 13 -

Xinhua News Agency

(End)

Beijing, February 13 -

Statistical Machine Translation Rules
Parsing and Word Alignment
• Information about parses can improve alignment quality (DeNero and Klein, 2007; Fossum et al, 2008; Pauls et al, 2010).
• Information about parses can improve alignment quality (DeNero and Klein, 2007; Fossum et al, 2008; Pauls et al, 2010).

• Information about alignment can improve parse quality (Smith and Smith, 2004; Burkett and Klein, 2008).
Parsing and Word Alignment

- Information about parses can improve alignment quality (DeNero and Klein, 2007; Fossum et al, 2008; Pauls et al, 2010).

- Information about alignment can improve parse quality (Smith and Smith, 2004; Burkett and Klein, 2008).

- This talk: performing both tasks jointly.
Weak Synchronization

High levels of product and project levels of high levels.
Weak Synchronization
Weak Synchronization

High levels of product and project of levels.
Weak Synchronization
Weak Synchronization

项目水平

产品

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高

水平

和

产品

project

and

product

of

levels

High

levels

和

产品

项目

项目和产品

的

高
Weak Synchronization

products and project levels of high
Weak Synchronization

High levels of product and project levels.
Weak Synchronization

High levels of product and project.
Weak Synchronization

High levels of product and project levels of high.
Weak Synchronization

High levels of product and project levels of product.
Parsing and Alignment Input

- project
- and
- product
- of
- levels
- High
Parsing and Alignment Input

Sentences

\((s, s')\)

- High
- levels
- project
- and
- product
- of
- 产品
- 项目
- 水平
- 高
Trees

project and product of levels High
High levels of product and project 音箱 levels 项目 水平 高
High levels of product and project levels of product and project 产品、项目 和水平 高
Trees contain Nodes

- **High levels** of product and project
- **n**
- **t**
- **t’**

This diagram illustrates the parsing and alignment output with nodes labeled with parts of speech (POS) such as **NP** (noun phrase), **PP** (prepositional phrase), **NN** (noun), **IN** (preposition or subordinating conjunction), **CC** (coordinating conjunction), **JJ** (adjective), **NNS** (noun phrase, plural), **VP** (verb phrase), **IP** (infinitival phrase), and **VA** (verbals). The alignment is shown between the English and Chinese text.
Trees contain Nodes

$t$ $t'$

$n$ $n'$

High levels of product and project 产品、项目

levels

of

product

高

水品
Alignments

project and product of levels

High
High levels of product and project PageSize Output

Alignments  α
Alignments contain Bispans
High levels of product and project.
High levels of product and project (t, a, t')
High levels of product and project.
Parsing and Alignment Output

(t, a, t')

NP  PP  NP  CC  NN

project and product of levels

High

产品

项目

水平

高
Parsening and Alignment Output

Tralignments

\((t, a, t')\)

Berkeley NLP
Loglinear Model

\[ P(t, a, t' | s, s') \propto \exp(\langle \theta, \phi(t, a, t') \rangle) \]
Loglinear Model

\[ P(t, a, t' | s, s') \propto \exp(\langle \theta, \phi(t, a, t') \rangle) \]

\[ \phi(t, a, t') \]
\[ \phi(n) \text{ TrgMono } = \log P_{\text{eng}}(n) \]
Monolingual Parsing Features

\[
\phi(n) \quad T_{\text{TrgMono}} = \log P_{\text{eng}}(\cdot)
\]

\[
\phi(n') \quad S_{\text{SrcMono}} = \log P_{\text{chi}}(\cdot)
\]
Alignment Features

\( \phi(b) \)

- High levels of product and project
- Project and product of levels
- 高水平
UnsupScore = \phi(b)

UnsupScore = P_{uns}(\text{product})
Alignment Features

\[
\text{UnsupScore} = \phi(b) \quad \text{LEX}\langle \text{product}, \text{产品} \rangle = 1
\]
High levels of product and project synchronization features.
Target-Alignment Synchronization Features
Target-Alignment Synchronization Features

\[ \phi(n, b) \]

\[ \text{TrgAlign}\langle \text{NP} \rangle = 1 \]
High levels of product and project .DataContext
Source-Alignment Synchronization Features

\[ \phi(b, n') \]

\[ \text{SrcAlign} \langle \text{NP} \rangle = 1 \]
Target-Source Synchronization Features

High levels of product and project.
Target-Source Synchronization Features

\[ \phi(n, b, n') \]

\[ \text{SrcTrg} \langle \text{NP}, \text{NP} \rangle = 1 \]
Factorized Feature Function

High levels of product and project
Approximate Viterbi Inference
Approximate Viterbi Inference

Given: $s, s'$
Approximate Viterbi Inference

Given: $s, s'$

Wanted: $\text{argmax } P(t, a, t'|s, s')$

$\text{argmax } P(t, a, t'|s, s')$
Approximate Viterbi Inference

Given: \( s, s' \)

Wanted: \( \arg\max_{(t,a,t')} P(t, a, t'| s, s') \)

Problem: Synchronization features
Approximate Viterbi Inference

Given: \( s, s' \)

Wanted: \( \text{argmax } P(t, a, t'| s, s') \)

Problem: Synchronization features

Solution:
Approximate Viterbi Inference

Given: \( s, s' \)

Wanted: \( \arg\max_{(t,a,t')} P(t,a,t'|s,s') \)

Problem: Synchronization features

Solution: 1. Fix \( a, t' \), find \( \arg\max_{t} P(t|a,t', s, s') \)
Approximate Viterbi Inference

Given: \( s, s' \)

Wanted: \( \arg\max_{(t,a,t')} P(t,a,t'|s,s') \)

Problem: Synchronization features

Solution: 1. Fix \( a, t' \), find \( \arg\max_t P(t|a,t',s,s') \)

2. Fix \( t, t' \), find \( \arg\max_a P(a|t,t',s,s') \)
Approximate Viterbi Inference

Given: \( s, s' \)

Wanted: \( \arg\max_{(t,a,t')} P(t,a,t'|s,s') \)

Problem: Synchronization features

Solution: 1. Fix \( a, t' \), find \( \arg\max_{t} P(t|a,t',s,s') \)
2. Fix \( t, t' \), find \( \arg\max_{a} P(a|t,t',s,s') \)
3. Fix \( t, a \), find \( \arg\max_{t'} P(t'|t,a,s,s') \)

Viterbi Inside Algorithm
Approximate Viterbi Inference

Given: $s, s'$

Wanted: $\arg\max_{(t,a,t')} P(t, a, t' | s, s')$

Problem: Synchronization features

Solution:

1. Fix $a, t'$, find $\arg\max_t P(t | a, t', s, s')$
2. Fix $t, t'$, find $\arg\max_a P(a | t, t', s, s')$
3. Fix $t, a$, find $\arg\max_{t'} P(t' | t, a, s, s')$
4. Repeat until convergence
Structured Mean Field Inference
Structured Mean Field Inference

Given: $s, s'$
Structured Mean Field Inference

Given: $s, s'$

Wanted: $P(t, a, t'|s, s')$
Structured Mean Field Inference

Given: $s, s'$

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$

$= q(t)q(a)q(t')$
Structured Mean Field Inference

Given: $s, s'$

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$

$= q(t)q(a)q(t')$

Algorithm:
Structured Mean Field Inference

Given: \( s, s' \)

Wanted: \( P(t, a, t' | s, s') \)

Approximation: \( P(t, a, t' | s, s') \approx Q(t, a, t') \)

\[ = q(t)q(a)q(t') \]

Algorithm: 1. \( q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle) \)
Structured Mean Field Inference

Given: $s, s'$

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t') = q(t)q(a)q(t')$

Algorithm: 1. $q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle)$

2. $q(a) \propto \exp(\langle \theta, \phi(E_q(t), a, E_q(t')) \rangle)$
Structured Mean Field Inference

Given: \( s, s' \)

Wanted: \( P(t, a, t' | s, s') \)

Approximation: \( P(t, a, t' | s, s') \approx Q(t, a, t') = q(t)q(a)q(t') \)

Algorithm: 1. \( q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle) \)

2. \( q(a) \propto \exp(\langle \theta, \phi(E_q(t), a, E_q(t')) \rangle) \)

3. \( q(t') \propto \exp(\langle \theta, \phi(E_q(t), E_q(a), t') \rangle) \)
Structured Mean Field Inference

Given: \( s, s' \)

Wanted: \( P(t, a, t' | s, s') \)

Approximation: \( P(t, a, t' | s, s') \approx Q(t, a, t') = q(t)q(a)q(t') \)

Algorithm:
1. \( q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle) \)
2. \( q(a) \propto \exp(\langle \theta, \phi(E_q(t), a, E_q(t')) \rangle) \)
3. \( q(t') \propto \exp(\langle \theta, \phi(E_q(t), E_q(a), t') \rangle) \)
4. Repeat until convergence
Training Data
Training Data

Parallel Treebank
Training Data

Parallel Treebank

Chinese Treebank
Training Data

Parallel Treebank

Chinese Treebank

English Treebank
Training Data

Parallel Treebank

Chinese Treebank

English Treebank

Bitext
Incomplete Training Data
Incomplete Training Data

Required data:

$$(t_i, a_i, t'_i)$$
Incomplete Training Data

Required data:

\( (t_i, a_i, t_i') \)

Actual data:

\( (t_i, w_i, t_i') \)
Incomplete Training Data
Incomplete Training Data

Word alignment $w$
Incomplete Training Data

Word alignment $w$

Legal alignments $a$
Parsing Improvement

86

84

82

80

Chinese

English

Monolingual

Joint
Parsing Improvement

- Chinese: Monolingual - 83.6, Joint - 85.7
- English: Joint - 85.7
Parsing Improvement

<table>
<thead>
<tr>
<th>Language</th>
<th>Monolingual</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>83.6</td>
<td>85.7</td>
</tr>
<tr>
<td>English</td>
<td>81.2</td>
<td>84.5</td>
</tr>
</tbody>
</table>
Incorrect English PP Attachment

... were established in such places as Beijing, Shanghai, etc.

在
北京
上海
等地
设立了...

...
Incorrect English PP Attachment

... were established in such places as Beijing, Shanghai, etc.

In Chinese:

在
北京
上海
等地
设立了...
were established in such places as Beijing, Shanghai, etc.
Parsing Improvement

Incorrect English PP Attachment

...were established in such places as Beijing, Shanghai, etc.

In Chinese:

在北京等地设立了...
Parsing Improvement

Incorrect English PP Attachment

...were established in such places as Beijing, Shanghai, etc.
were established in such places as Beijing, Shanghai, etc.
were established in such places as Beijing, Shanghai etc.
... were established in such places as Beijing, Shanghai, etc.
<table>
<thead>
<tr>
<th>Alignment F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
</tr>
<tr>
<td>85</td>
</tr>
<tr>
<td>80</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>65</td>
</tr>
<tr>
<td>Alignment F1</td>
</tr>
</tbody>
</table>
Word Alignment Improvement

<table>
<thead>
<tr>
<th>HMM</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>69.5</td>
<td>85.0</td>
</tr>
</tbody>
</table>

Alignment F1
Word Alignment Improvement

Alignment F1

- HMM: 69.5
- Supervised ITG: 79.5
- Joint: 85.0
Word Alignment Improvement

- **Alignment F1**
  - HMM: 69.5
  - Supervised ITG: 79.5
  - Joint: 85.0

- **BLEU (Hiero)**
  - HMM: 29.4
  - Supervised ITG: 29.4
  - Joint: 30.6
Word Alignment Improvement

<table>
<thead>
<tr>
<th>Method</th>
<th>Alignment F1</th>
<th>BLEU (Hier)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>69.5</td>
<td>29.4</td>
</tr>
<tr>
<td>Supervised ITG</td>
<td>79.5</td>
<td>30.4</td>
</tr>
<tr>
<td>Joint</td>
<td>85.0</td>
<td>30.6</td>
</tr>
</tbody>
</table>
Currently, the cause of the plane crash is still not clear.

Reference
At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)
The cause of planes is still not clear yet.
Currently the cause of the plane crash is still unclear.

Reference
At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)
The cause of planes is still not clear yet.
Reference
At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)
The cause of planes is still not clear yet.
Currently, the cause of the plane crash is still unclear.

Reference
At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)
The cause of planes is still not clear yet.
Translation Improvement

Reference
At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)
The cause of planes is still not clear yet.

Joint Model
The cause of plane collision remained unclear.
Conclusions
Conclusions

• Alignments and syntax are mutually constraining.
Conclusions

• Alignments and syntax are mutually constraining.

• The tasks of parsing and word alignment both gain from joint, weakly synchronized modeling.
Conclusions

- Alignments and syntax are mutually constraining.

- The tasks of parsing and word alignment both gain from joint, weakly synchronized modeling.

- Our bitext parse-and-aligner will be available for download at http://nlp.cs.berkeley.edu.
Thank You!