Learning Joint Semantic Parsers from Disjoint Data

Hao Peng\textsuperscript{1}, Sam Thomson\textsuperscript{2}, Swabha Swayamdipta\textsuperscript{2}, Noah A. Smith\textsuperscript{1}

\textsuperscript{1}University of Washington
\textsuperscript{2}Carnegie Mellon University

@NAACL
June 4, 2018
Motivations

❖ Larger data \(\xrightarrow{\text{almost}}\) Better performance

❖ Overlaps among different theories
Overview

Learning Joint Semantic Parsers from Disjoint Data

FrameNet vs. semantic dependencies
Different structures; no parallel annotations
Overview

Learning Joint Semantic Parsers from Disjoint Data

Joint decoding
Latent variables

FrameNet vs. semantic dependencies
Different structures; no parallel annotations
Outline

- Parsing semantic spans and dependencies
- Joint parsing
- Learning with latent variables
- Empirical results
Input:

A few books fell in the room.

Target: token span

Lexical unit: lemma.pos

Baker et al., (1998)
Parsing FrameNet Structures

Input:

A few books fell in the room .

Target: token span

Lexical unit: lemma.pos

Output:

Frame

A few books fell in the room .

Arguments: span + semantic roles

Baker et al., (1998)
Parsing FrameNet Structures

Input:

A few books fell in the room.

Score:

\[ F (\text{A few books} \quad \text{fell} \quad \text{in the room}. ) \]

\[ f_{\text{frame}} (\text{Motion Directional}) + f_{\text{arg}} (\text{Theme}) + f_{\text{arg}} (\text{Place}) \]
A few books fell in the room.

Score:

\[ F \left( \text{A few books \textcolor{red}{fell} in the room} \right) \]

\[ f_{\text{frame}}(\text{Motion Directional}) + f_{\text{arg}}(\text{Theme}) + f_{\text{arg}}(\text{Place}) \]

BiLSTM+MLPs
Parsing FrameNet Structures

Decoding:

\[
\max_{\text{frame, args}} F \left( \begin{array}{c}
\text{A few books fell in the room .}
\end{array} \right)
\]

\text{s.t.} \quad \bullet \text{non-overlapping} \quad \bullet \text{consistency} \quad \bullet \ldots

Dynamic program
Kong et al., (2016); Swayamdipta et al., (2017)
Parsing Semantic Dependencies

Input:
A few books fell in the room.

Output:
MRS-derived dependencies (DM)

Oepen et al., (2015)
Parsing Semantic Dependencies

Input:
A few books fell in the room.

Score:
\[ G \left( \begin{array}{cccc}
\text{mwe} & \text{arg1} & \text{arg1} & \text{arg1} \\
\text{A few books} & \text{fell} & \text{in} & \text{the room} \\
\end{array} \right) \]

\[ \sum g \left( \begin{array}{ccc}
\text{head} & \text{mod} \\
labeled arcs \\
\end{array} \right) \text{BiLSTM+MLPs} \]
Parsing Semantic Dependencies

Decoding:

Linear program
\[ \text{AD}^3; \text{Martins et al., (2011)} \]

\[ \max \]
[\( \text{labeled arcs} \)]

\[ G \left( \begin{array}{c}
\text{few books} \\
\text{arg1} \\
\text{fell} \\
\text{room} \\
\text{...} \\
\end{array} \right) \]

s.t.
- consistency
- determinism
- ...
Outline

- Parsing semantic spans and dependencies
- Joint parsing
- Learning with latent variables
- Empirical results
Joint Parsing

Sharing parameters:
Swayamdipta et al., (2016); Hershcovich et al., (2018)

\[
F \left( \frac{A \ few \ books \ \text{fell} \ in \ the \ room.}{\text{Theme} \ \text{Motion} \ \text{Place}} \right) \quad G \left( \frac{A \ few \ books \ fell \ in \ the \ room.}{\text{mwe} \ \text{arg1} \ \text{arg1} \ \text{arg1} \ \text{BV} \ \text{top} \ \text{arg2}} \right)
\]

Shared LSTMs
Joint Parsing

Sharing parameters:
Swayamdipta et al., (2016); Hershcovich et al., (2018)

\[
F \begin{pmatrix}
A \text{ few books} & \text{fell in the room} \\
\text{Theme} & \text{Motion Directional} & \text{Place}
\end{pmatrix}
\]

\[
G \begin{pmatrix}
A \text{ few books fell in the room} \\
\text{arg2} & \text{arg1} & \text{arg1} & \text{arg1} & \text{BV}
\end{pmatrix}
\]

Shared LSTMs

This work, joint decoding:

\[
H \begin{pmatrix}
A \text{ few books} & \text{fell in the room} \\
\text{Theme} & \text{Motion Directional} & \text{Place}
\end{pmatrix}
\]
Joint Parsing

Sharing parameters:
Swayamdipta et al., (2016); Hershcovich et al., (2018)

\[
F \left( \frac{\text{A few books fell in the room}.}{} \right) \quad G \left( \frac{\text{A few books fell in the room}.}{} \right)
\]

Shared LSTMs

This work, joint decoding:

\[
H \left( \frac{\text{A few books fell in the room}.}{} \right)
\]

Orthogonal
Joint Parsing

Input:  
A few books *fell* in the room .  

*fall.*

Score:  

\[ H \left( \begin{array}{c}
A \text{ few books } \textbf{fell} \text{ in the room .}
\end{array} \right) \]

*Theme*  
*Motion*  
*Directional*  
*Place*
Joint Parsing

Input:
A few books fell in the room.

Score:

\[
H \left( \begin{array}{c}
\text{A few books} \\
\text{fell} \\
\text{in the room}.
\end{array} \right)
\]

\[
F \left( \begin{array}{c}
\text{A few books} \\
\text{fell} \\
\text{in the room}.
\end{array} \right) + G \left( \begin{array}{c}
\text{A few books} \\
\text{fell} \\
\text{in the room}.
\end{array} \right)
\]

FrameNet Score

DM Score
Joint Parsing

Input: A few books fell in the room.

Score:

\[ F(H) + G(H) + h_{\text{joint}}(?) \]

FrameNet Score
Affinities between them
DM Score

\[ H \]
A few books fell in the room.

\( \text{fall.v} \)

Theme
Motion
Place
arg1
arg1
arg1
arg2
mwe
arg1
arg1
BV
top

\[ F \]
A few books fell in the room.

\( \text{fall.v} \)

Theme
Motion
Place
arg1
arg1
arg1
arg2
mwe
arg1
arg1
BV
top

\[ G \]
A few books fell in the room.
Span vs. Dependencies

$ h_{\text{joint}}(?)$

If both were dependencies
Lluís et al., (2013); Peng et al., (2017)

$ h_{\text{joint}}(\text{head mod})$

If both were spans
Finkel and Manning, (2009)

$ h_{\text{joint}}(\text{role1 role2})$
Span vs. Dependencies

\[ h_{\text{joint}}(?) \]

If both were dependencies
Lluís et al., (2013); Peng et al., (2017)
\[ h_{\text{joint}}(\text{head mod}) \]

If both were spans
Finkel and Manning, (2009)
\[ h_{\text{joint}}(\text{role1 role2}) \]

Structural divergence
A few books fell
\[ \text{mwe} \quad \text{arg1} \quad \text{arg1} \]
\[ \text{Theme} \quad \text{fall. v} \quad \text{Motion Directional} \]
Span vs. Dependencies

Structural divergence

Designate a head for each span
PropBank dependencies; Surdeanu et al., (2008)
Span vs. Dependencies

Structural divergence

Designate a head for each span
PropBank dependencies; Surdeanu et al., (2008)

Head selected by syntax
Span vs. Dependencies

Structural divergence

Designate a head for each span
PropBank dependencies; Surdeanu et al., (2008)
Span vs. Dependencies

Structural divergence

This work
Span vs. Dependencies

Score:

\[ H \left( \text{A few books fell in the room} \right) \]  
\[ F \left( \text{A few books fell in the room} \right) + G \left( \text{A few books fell in the room} \right) \]  
\[ + h_{\text{joint}} \left( \text{A few books fell in the room} \right) \]

FrameNet Score

Affinities between them

Multilinear mapping

DM Score
Span vs. Dependencies

Decoding:

\[
\max_H \left( \begin{array}{c}
\text{frame, args} \\
\text{labeled arcs} \\
\text{joint parts}
\end{array} \right)
\]

Linear program
Speed up by promoting sparsity
Outline

- Parsing semantic spans and dependencies
- Joint parsing
- Learning with latent variables
- Empirical results
Learning with Latent Variables

FrameNet data

DM data
Learning with Latent Variables

FrameNet data

DM data

Supervision

Theme

role

head mod

A few books fell

Theme

fall.v

Supervision

Theme

role

head mod

A few books fell

Theme

fall.v
Learning with Latent Variables

Latent structured hinge
Yu and Joachims, (2009)

\[ \mathcal{L} = - \max H(A \text{ few books } \text{fell}_v \text{ in the room}) \]

\[ + \max H(A \text{ few books } \text{fell}_v \text{ in the room}) + \delta \]

FrameNet data
Learning with Latent Variables

Latent structured hinge
Yu and Joachims, (2009)

\[ \mathcal{L} = - \max H (\text{A few books} \xrightarrow{\text{fell}} \text{in the room}) \]

\[ + \max H (\text{A few books} \xrightarrow{\text{fell,v}} \text{in the room}) \]

FrameNet data

Prediction

cost
Learning with Latent Variables

Latent structured hinge
Yu and Joachims, (2009)

\[ \mathcal{L} = -\max H + \max H \]

Gold FN output

FrameNet data
Learning with Latent Variables

Latent structured hinge
Yu and Joachims, (2009)

$$\mathcal{L} = - \max H \left( \begin{array}{c}
aligned{\text{A few books} & & \text{fell} & & \text{in} & & \text{the} & & \text{room} \end{array} \right) \right) + \max H \left( \begin{array}{c}
aligned{\text{arg1} & & \text{arg2} & & \text{arg3} & & \text{BV} \end{array} \right) \right) + \delta$$

FrameNet data
Outline

- Parsing semantic spans and dependencies
- Joint parsing
- Learning with latent variables
- Empirical results
FrameNet Results

Compared models:

<table>
<thead>
<tr>
<th>FitzGerald et al. (2015)</th>
<th>Open-SESAME</th>
<th>Yang &amp; Mitchell (2017)</th>
<th>This work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame &amp; Arg. ID.</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

Open-SESAME: Swayamdipta et al., (2017)
FrameNet Results

Compared models:

- FitzGerald et al. (2015)
- Open-SESAME
- Yang & Mitchell (2017)
- This work

Frame & Arg. ID.

Pipeline

Predict both frames and arguments

Open-SESAME: Swayamdipta et al., (2017)
FrameNet Results

Compared models:

- FitzGerald et al. (2015)
- Open-SESAME
- Yang & Mitchell (2017)
- This work

FrameNet & PropBank

- FrameNet & PropBank
- FrameNet & Syntax
- FrameNet & DM

Frame & Arg. ID.

Multitask Learning

Open-SESAME: Swayamdipta et al., (2017)
FrameNet Results

Compared models:

- FitzGerald et al. (2015)
- Open-SESAME
- Yang & Mitchell (2017)
- This work

<table>
<thead>
<tr>
<th>Feature</th>
<th>FitzGerald et al. (2015)</th>
<th>Open-SESAME</th>
<th>Yang &amp; Mitchell (2017)</th>
<th>This work</th>
<th>Basic</th>
<th>w/o joint decoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame &amp; Arg. ID.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multitask Learning</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Joint Decoding</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Open-SESAME: Swayamdipta et al., (2017)
### FrameNet Results

**Frame and argument F$_1$, FrameNet 1.5 test set**

<table>
<thead>
<tr>
<th>Method</th>
<th>F$_1$</th>
<th>Single</th>
</tr>
</thead>
<tbody>
<tr>
<td>FitzGerald et al. (2015)</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Open-SESAME</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Yang &amp; Mitchell (2017)</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>This work</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

- **Frame & Arg. ID.**
  - Checkmark indicates the method uses this technique.

- **Multitask Learning**
  - Checkmark indicates the method uses this technique.

- **Joint Decoding**
  - Checkmark indicates the method uses this technique.

- **Open-SESAME**: Swayamdipta et al., (2017)
### FrameNet Results

Frame and argument $F_1$, FrameNet 1.5 test set

<table>
<thead>
<tr>
<th></th>
<th>single</th>
<th>ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitzgerald et al. (2015)</td>
<td>60</td>
<td>64</td>
</tr>
<tr>
<td>Open-SESAME</td>
<td>72</td>
<td>76</td>
</tr>
<tr>
<td>Yang &amp; Mitchell (2017)</td>
<td>5×</td>
<td>2×</td>
</tr>
<tr>
<td>This work</td>
<td>10×</td>
<td>2××</td>
</tr>
</tbody>
</table>

- **Frame & Arg. ID.**: ✔
- **Multitask Learning**: ✔
- **Joint Decoding**: ✔

Open-SESAME: Swayamdipta et al., (2017)
FrameNet Results

Frame and argument $F_1$, FrameNet 1.5 test set

<table>
<thead>
<tr>
<th>Method</th>
<th>Single</th>
<th>Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>FitzGerald et al. (2015)</td>
<td>60</td>
<td>72</td>
</tr>
<tr>
<td>Open-SESAME</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>Yang &amp; Mitchell (2017)</td>
<td>68</td>
<td>72$^*$</td>
</tr>
<tr>
<td>This work</td>
<td>10$^*$</td>
<td>80</td>
</tr>
<tr>
<td>Basic</td>
<td></td>
<td>2$^*$</td>
</tr>
<tr>
<td>w/o joint decoding</td>
<td></td>
<td>2$^*$</td>
</tr>
</tbody>
</table>

$^*$ Open-SESAME: Swayamdipta et al., (2017)

- **Frame & Arg. ID.**
  - **Multitask Learning**
  - **Joint Decoding**

The table shows the $F_1$ scores for FrameNet 1.5 test set with different methods and configurations. The $F_1$ scores are marked with checks in the corresponding columns for each method.
Frame ID. accuracy, FrameNet 1.5 test set

Accuracy

92
90
88
86
84
82

Hartmann  Y & M  Hermman  This work
Frame ID. accuracy, FrameNet 1.5 test set

DM labeled $F_1$ SemEval ’15 test set

Conclusion

Problem

A few books fell in the room.

FrameNet data

DM data
Conclusion

Problem

A few books fell in the room.

FrameNet data

DM data

Method

\[ \mathcal{L} = \max H \left( \text{A few books} \, \text{fell} \, \text{in the room} \right) + \delta \]
Conclusion

Problem

A few books fell in the room.

Theme: mwe, arg1, arg1, top, arg1
Motion: fall_v, Directional
Place: arg2, BV

FrameNet data  DM data

Method

\[ \mathcal{L} = - \max_{\text{joint parts}} \mathcal{H} \left( \frac{\text{A few books fall_v in the room}}{\text{arg1, arg2, BV}} \right) + \max_{\text{frame, args joint parts}} \mathcal{H} \left( \frac{\text{A few books fall_v in the room}}{\text{arg1, BV}} \right) + \delta \]

Results

![Graph showing results with single and ensemble models]
I thank you

Communicator: arg1
Judgement: thank.v
Direct Address: arg2
Addressee: you