A Simple and Accurate Syntax-Agnostic Neural Model for Dependency-based Semantic Role Labeling

Diego Marcheggiani, Anton Frolov, and Ivan Titov

University of Amsterdam
Yandex
University of Edinburgh

CoNLL 2017, Vancouver
Contributions

- Neural model for dependency-based SRL
- Simple
- Syntax-agnostic
- State of the art on out-of-domain data
- State-of-the-art performance on English, Chinese, Czech, and Spanish
Semantic role labeling

- Predicting the predicate-argument structure of a sentence

Sequa makes and repairs jet engines.
Semantic role labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates

Sequa makes and repairs jet engines.
Semantic role labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles

Sequa makes and repairs jet engines.
Semantic role labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles

Sequa makes and repairs jet engines.
Semantic role labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles

Sequa makes and repairs jet engines.

- make.01
- repair.01
- engine.01
- A0
- A1
Semantic role labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles
Semantic role labeling

- Only the head of an argument is labeled
- Sequence labeling task for each predicate
- Focus on argument identification and labeling

Sequa makes and repairs jet engines.
Model Architecture

- Word representation
- Sentence encoding (BiLSTM)
- Local classifier
Word encoding

Lane disputed those estimates
Word encoding

- pretrained word embeddings
Word encoding

- pretrained word embeddings
- word embeddings
Word encoding

- pretrained word embeddings
- word embeddings
- POS tag embeddings
Word encoding

- pretrained word embeddings
- word embeddings
- POS tag embeddings
- predicate lemmas embeddings
Sentence encoding

- **Bidirectional LSTM**
  - Forward LSTM encodes left context
  - Backward LSTM encodes right context
  - Forw. and Backw. states are concatenated
  - Stacking of several BiLSTM layers
Local classifier

Lane disputed those estimates
Local classifier

\[ p(r|v_i, p) \propto \exp(W_r v_i) \]
Local classifier

\[ p(r|v_i, p) \propto \exp(W_r v_i) \]
Local classifier

\[ p(r|v_i, p) \propto \exp(W_r v_i) \]
Re-encoding

Encoding predicates

- Add a predicate flag to word representation
- For each predicate the sentence is re-encoded

\[ p(r|v_i, p) \propto \exp(W_r v_i) \]
Encoding predicates

- Add a predicate flag to word representation
- For each predicate the sentence is re-encoded

\[ p(r|v_i, p) \propto \exp(W_r v_i) \]

Zhou and Xu, (2015)
Compositional classifier

Lane disputed those estimates
Compositional classifier

\[ p(r|t_i, t_p, l) \propto \exp(W_{l,r}(t_i \circ t_p)) \]
Compositional classifier

Kiperwasser and Goldberg, (2016)

\[ p(r|t_i, t_p, l) \propto \exp(W_{l,r}(t_i \circ t_p)) \]

predicate representation

candidate argument representation

A1

Local classifier

K layers BiLSTM

word representation

Lane disputed those estimates
Compositional classifier

\[ p(r|t_i, t_p, l) \propto \exp(W_{l,r}(t_i \circ t_p)) \]

\[ W_{l,r} = ReLU(U(q_l \circ q_r)) \]

Kiperwasser and Goldberg, (2016)
Compositional classifier

\[ p(r|t_i, t_p, l) \propto \exp(W_{l,r}(t_i \odot t_p)) \]

\[ W_{l,r} = \text{ReLU}(U(q_l \odot q_r)) \]

Fitzgerald et al., (2015)

Kiperwasser and Goldberg, (2016)
Experimental setting

- CoNLL-2009 dependency-based SRL dataset (standard split)
  - English, Chinese, Czech, Spanish
  - F1 as evaluation measure
- State-of-the-art predicate disambiguation models
- Hyperparameters tuned on English dev set
- Adam optimizer
Predicate encoding

CoNLL 2009 development

Compositional classifier: 80.4

Re-encoding

Both
Predicate encoding

CoNLL 2009 development

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compositional classifier</td>
<td>80.4</td>
</tr>
<tr>
<td>Re-encoding</td>
<td>85.6</td>
</tr>
<tr>
<td>Both</td>
<td></td>
</tr>
</tbody>
</table>
Predicate encoding

CoNLL 2009 development

- Compositional classifier: 80.4
- Re-encoding: 85.6
- Both: 86.6
English Test set

Björkelund et al. (2010) (global): 86.9
FitzGerald et al. (2015) (global): 87.3
Roth and Lapata (2016) (local): 86.7
Roth and Lapata (2016) (global): 87.7
Ours (local): 87.7
English Test set

Björkelund et al. (2010) (global) - 86.9
FitzGerald et al. (2015) (global) - 87.3
Roth and Lapata (2016) (local) - 86.7
Roth and Lapata (2016) (global) - 87.7
Ours (local) - 87.7
English Test set

86.9

87.3

86.7

87.7

Björkelund et al. (2010) (global)

FitzGerald et al. (2015) (global)

Roth and Lapata (2016) (local)

Roth and Lapata (2016) (global)

Ours (local)

On par with global syntax-rich models
English out-of-domain

- Björkelund et al. (2010) (global) - 75.7
- FitzGerald et al. (2015) (global) - 75.2
- Roth and Lapata (2016) (local) - 75.3
- Roth and Lapata (2016) (global) - 76.1
- Ours (local)
English out-of-domain

Björkelund et al. (2010) (global) 75.7
FitzGerald et al. (2015) (global) 75.2
Roth and Lapata (2016) (local) 75.3
Roth and Lapata (2016) (global) 76.1
Ours (local) 77.7
English out-of-domain

Björkelund et al. (2010) (global) 75.7
FitzGerald et al. (2015) (global) 75.2
Roth and Lapata (2016) (local) 75.3
Roth and Lapata (2016) (global) 76.1
Ours (local) 77.7

1.6%
Björkelund et al. (2010) (global) 75.7
FitzGerald et al. (2015) (global) 75.2
Roth and Lapata (2016) (local) 75.3
Roth and Lapata (2016) (global) 76.1
Ours (local) 77.7

Robust on out-of-domain data
Chinese and Spanish Test set

**Chinese Test results**

- Zhao et al. (2009) (global): 77.7
- Björkelund et al. (2009) (global): 78.6
- Roth and Lapata (2016) (global): 79.4
- Ours (local): 81.2

**Spanish Test results**

- Zhao et al. (2009) (global): 80.5
- Björkelund et al. (2009) (global): 76.5
- Roth and Lapata (2016) (global): 80.2
- Ours (local): 80.3
Czech test set and out-of-domain

Czech Test results

Zhao et al. (2009) (global): 85.2%
Björkelund et al. (2009) (global): 85.4%
Ours (local): 86%

Czech Out-of-domain results

Zhao et al. (2009) (global): 85.4%
Björkelund et al. (2009) (global): 83.9%
Ours (local): 87.2%
Czech test set and out-of-domain

Czech Test results

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhao et al. (2009) (global)</td>
<td>85,2</td>
</tr>
<tr>
<td>Björkelund et al. (2009) (global)</td>
<td>85,4</td>
</tr>
<tr>
<td>Ours (local)</td>
<td>86</td>
</tr>
</tbody>
</table>

Czech Out-of-domain results

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhao et al. (2009) (global)</td>
<td>85,4</td>
</tr>
<tr>
<td>Björkelund et al. (2009) (global)</td>
<td>83,9</td>
</tr>
<tr>
<td>Ours (local)</td>
<td>87,2</td>
</tr>
</tbody>
</table>

1.8% improvement in out-of-domain results
Czech test set and out-of-domain

Czech Test results

Zhao et al. (2009) (global): 85.2
Björkelund et al. (2009) (global): 85.4
Ours (local): 86

Czech Out-of-domain results

Zhao et al. (2009) (global): 85.4
Björkelund et al. (2009) (global): 83.9
Ours (local): 87.2

1.8%
Distance analysis
Distance analysis

Long-range dependencies are better captured
Conclusion

- Simple syntax-agnostic dependency-based SRL model
- Very robust on out-of-domain data
- Building block for syntax-aware models (Graph convolutional nets)
  - Marcheggiani and Titov, EMNLP 2017
Conclusion

- Simple syntax-agnostic dependency-based SRL model
- Very robust on out-of-domain data
- Building block for syntax-aware models (Graph convolutional nets)
  - Marcheggiani and Titov, EMNLP 2017
  - [github.com/diegma/neural-dep-srl](https://github.com/diegma/neural-dep-srl)

**Funding:**
- ERC StG BroadSem 678254
- NWO VIDI 639.022.518
- Amazon Web Services (AWS) grant