Learning Cross-lingual Distributed Logical Representations for Semantic Parsing

Yanyan Zou and Wei Lu  
StatNLP Group  
Singapore University of Technology and Design  
July 18, 2018
Outline

✓ Background & Motivation
✓ Method
✓ Experiments & Analysis
✓ Conclusion
Semantic Parsing

Goal: Map natural languages into semantic representations.
Semantic Parsing

Goal: Map natural languages into semantic representations.

English: what states have no bordering state?
Semantic Parsing

Goal: Map natural languages into semantic representations.

Natural Language

English: what states have no bordering state?

Logical Form

answer(exclude(state(all), next_to(state(all)))))
Semantic Parsing

Goal: Map natural languages into semantic representations.

**Natural Language**

English: what states have no bordering state?

**Semantic Tree**

\[
\text{QUERY: answer (STATE)} \\
\text{STATE: exclude (STATE, STATE)} \\
\text{STATE: state (all) } \text{ STATE: next\_to (STATE)} \\
\text{ STATE: state (all)}
\]

**Logical Form**

\[
\text{answer(exclude(state(all), next\_to(state(all))))}
\]
Joint Representations

Proposed in previous works:

✓ Synchronous CFG derivation trees
  Wong and Mooney (2006, 2007)
✓ CCG derivation trees
  Zettlemoyer and Collins (2005, 2007)
✓ Bayesian tree transducers
  Jones, Goldwater and Johnson (2012)
✓ Hybrid Trees
  Lu, Ng, Lee, Zettlemoyer (2008)
Hybrid Tree

Input: what states have no bordering states?
Hybrid Tree

Input: what states have no bordering states?

QUERY: answer (STATE)

STATE: exclude (STATE, STATE)

STATE: state (all) have no bordering STATE: state (all)

what states

bordering states
Input: what states have no bordering states?

```
QUERY : answer (STATE)

STATE: exclude (STATE, STATE)  

STATE : state (all)  have no  STATE : next_to (STATE)  

what states  

bordering  

STATE : state (all)  

states
```

Output: \( \text{answer(exclude(state(all), next_to(state(all))))} \)
**Input:** what states have no bordering states?

**QUERY:** answer (STATE)

**STATE:** exclude (STATE, STATE)

STATE: state (all) have no

STATE: next_to (STATE)

what states

bordering

**STATE:** state (all)

states

\[ p(m, n) = \sum_{h \in H(n, m)} p(m, h, n) \]
Input: what states have no bordering states?

\[ p(m|n) = \sum_{h \in H(n,m)} p(m, h|n) \]
Neural Hybrid Tree

Input: what states have no bordering states?

QUERY: answer (STATE)

STATE: exclude (STATE, STATE)

STATE: state (all) have no

STATE: next_to (STATE)

what states

bordering

STATE: state (all)

states

- Neural hybrid tree is an extension of discriminative hybrid tree.
Neural Hybrid Tree

Input: what states have no bordering states?

QUERY: answer (STATE)

STATE: exclude (STATE, STATE)

STATE: state (all) have no what states

STATE: next_to (STATE)

bordering STATE: state (all) states
Neural Hybrid Tree

Output layer

Hidden layer

Input layer

Score vector

Discriminative hybrid tree

STATE : next_to (STATE)

bordering

STATE : state (all)

states

(Susanto and Lu, 2017)
Neural Hybrid Tree

Score vector

Output layer

Hidden layer

Input layer

Discriminative hybrid tree

STATE: next_to (STATE)

bordering

STATE: state (all)

Word window in size of (2J+1)

(states)
What do we have?

- English Sentences
- Semantic Trees
- Semantic Parser For English
What do we have?

- English Sentences
- Semantic Trees
- German Sentences
- Indonesian Sentences
- Chinese Sentences

Semantic Parser For English
What do we have?

<table>
<thead>
<tr>
<th>Background</th>
<th>Method</th>
<th>Experiments &amp; Analysis</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

- English Sentences
- Semantic Trees
- German Sentences
- Indonesian Sentences
- Chinese Sentences

Can we leverage multi-lingual resources to improve the performance of a monolingual semantic parser?
What do we have?

- English Sentences
- Semantic Trees
- German Sentences
- Indonesian Sentences
- Chinese Sentences

Can we leverage multi-lingual resources to improve the performance of a monolingual semantic parser?

The answer is Yes!!!
Setup

Target Language (E.g., English)

Semantic Trees

Semantic Parser
For English

Auxiliary Languages
German
Indonesian
Chinese

...
Setup

<table>
<thead>
<tr>
<th>Background</th>
<th>Method</th>
<th>Experiments &amp; Analysis</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

Target Language (E.g., English)

Semantic Trees

Auxiliary Languages
- German
- Indonesian
- Chinese
- ...

Semantic Parser For English

Cross-lingual information
Setups

Target Language (E.g., English)

Semantic Trees

Auxiliary Languages
  German
  Indonesian
  Chinese
  ...

Semantic Parser For English

We learn distributed representations of semantic units where such cross-lingual information is captured.
We learn distributed representations of semantic units where such cross-lingual information is captured.
Cross-lingual Representations

We construct a semantics-word co-occurrence matrix $C \in \mathbb{R}^{m \times n}$ based on auxiliary languages and semantic trees.
Cross-lingual Representations

The singular value decomposition (SVD) is then applied to the co-occurrence matrix, leading to

\[ C = U\Sigma V^* \]

We truncate the diagonal matrix \( \Sigma \) and left multiply it with \( U \) :

\[ R = U\tilde{\Sigma} \]
The singular value decomposition (SVD) is then applied to the co-occurrence matrix, leading to

\[ C = U \Sigma V^* \]

We truncate the diagonal matrix \( \Sigma \) and left multiply it with \( U \):

\[ R = U \tilde{\Sigma} \]

Each row in \( R \) is a d-dimensional vector, giving a low-dimensional representation for one semantic unit.
The singular value decomposition (SVD) is then applied to the co-occurrence matrix, leading to

\[ C = U \Sigma V^* \]

We truncate the diagonal matrix \( \Sigma \) and left multiply it with \( U \):

\[ R = U \tilde{\Sigma} \]

Each row in \( R \) is a d-dimensional vector, giving a low-dimensional representation for one semantic unit.

The learned representations are considered as features for discriminative and neural hybrid tree models.
## Results

**Data:** Multilingual Geoquery

<table>
<thead>
<tr>
<th>Method</th>
<th>HT-D (+O) J=0</th>
<th>HT-D (+O) J=1</th>
<th>HT-D (+O) J=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>86.1</td>
<td>86.1</td>
<td>89.6</td>
</tr>
<tr>
<td>Thai</td>
<td>86.1</td>
<td>86.1</td>
<td>86.1</td>
</tr>
<tr>
<td>German</td>
<td>81.1</td>
<td>84.6</td>
<td>90.0</td>
</tr>
<tr>
<td>Greek</td>
<td>73.9</td>
<td>76.8</td>
<td>83.9</td>
</tr>
<tr>
<td>Chinese</td>
<td>77.9</td>
<td>82.1</td>
<td>83.9</td>
</tr>
<tr>
<td>Indonesian</td>
<td>77.9</td>
<td>82.1</td>
<td>83.9</td>
</tr>
<tr>
<td>Swedish</td>
<td>75.4</td>
<td>78.6</td>
<td>74.6</td>
</tr>
<tr>
<td>Farsi</td>
<td>74.6</td>
<td>76.1</td>
<td>75.7</td>
</tr>
</tbody>
</table>

5 out of 8 languages get improved
Results without Neural Features

Data: Multilingual Geoquery
Baselines: (Lu et al., 2008) (Lu, 2015)

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Thai</th>
<th>German</th>
<th>Greek</th>
<th>Chinese</th>
<th>Indonesian</th>
<th>Swedish</th>
<th>Farsi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
</tr>
<tr>
<td>HT-G</td>
<td>76.8</td>
<td>81.0</td>
<td>73.6</td>
<td>76.7</td>
<td>62.1</td>
<td>68.5</td>
<td>56.1</td>
<td>58.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>69.3</td>
<td>74.6</td>
<td>66.4</td>
<td>72.8</td>
<td>61.4</td>
<td>70.5</td>
</tr>
<tr>
<td></td>
<td>51.8</td>
<td>58.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT-D</td>
<td>86.8</td>
<td>86.8</td>
<td>80.7</td>
<td>80.7</td>
<td>75.7</td>
<td>75.7</td>
<td>79.3</td>
<td>79.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>79.3</td>
<td>79.3</td>
<td>76.1</td>
<td>76.1</td>
<td>79.3</td>
<td>79.3</td>
</tr>
<tr>
<td></td>
<td>73.9</td>
<td>73.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Background | Method | Experiments & Analysis | Conclusion
Results without Neural Features

Data: Multilingual Geoquery
Baselines: (Lu et al., 2008) (Lu, 2015)
(+o): models with distributed representations of semantic units.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Thai</th>
<th>German</th>
<th>Greek</th>
<th>Chinese</th>
<th>Indonesian</th>
<th>Swedish</th>
<th>Farsi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
</tr>
<tr>
<td>HT-G</td>
<td>76.8</td>
<td>81.0</td>
<td>73.6</td>
<td>76.7</td>
<td>62.1</td>
<td>68.5</td>
<td>69.3</td>
<td>74.6</td>
</tr>
<tr>
<td>HT-D</td>
<td><strong>86.8</strong></td>
<td><strong>86.8</strong></td>
<td><strong>80.7</strong></td>
<td><strong>80.7</strong></td>
<td><strong>75.7</strong></td>
<td><strong>75.7</strong></td>
<td><strong>79.3</strong></td>
<td><strong>79.3</strong></td>
</tr>
<tr>
<td>HT-D (+O)</td>
<td>86.1</td>
<td>86.1</td>
<td><strong>81.1</strong></td>
<td><strong>81.1</strong></td>
<td>73.6</td>
<td>73.6</td>
<td>81.4</td>
<td>81.4</td>
</tr>
</tbody>
</table>
Results with Neural Features

Data: Multilingual Geoquery
Baselines: *(Lu et al., 2008) (Lu, 2015) (Susanto and Lu, 2017)*
(+o): models with distributed representations of semantic units.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HT-G</td>
<td>76.8</td>
<td>81.0</td>
<td>73.6</td>
<td>76.7</td>
<td>62.1</td>
<td>68.5</td>
<td>69.3</td>
<td>74.6</td>
<td>56.1</td>
<td>58.4</td>
<td>66.4</td>
<td>72.8</td>
<td>61.4</td>
<td>70.5</td>
<td>51.8</td>
<td>58.6</td>
</tr>
<tr>
<td>HT-D</td>
<td><strong>86.8</strong></td>
<td><strong>86.8</strong></td>
<td>80.7</td>
<td>80.7</td>
<td><strong>75.7</strong></td>
<td><strong>75.7</strong></td>
<td>79.3</td>
<td>79.3</td>
<td>76.1</td>
<td>76.1</td>
<td>75.0</td>
<td>75.0</td>
<td><strong>79.3</strong></td>
<td><strong>79.3</strong></td>
<td>73.9</td>
<td>73.9</td>
</tr>
<tr>
<td>HT-D (+O)</td>
<td>86.1</td>
<td>86.1</td>
<td><strong>81.1</strong></td>
<td><strong>81.1</strong></td>
<td>73.6</td>
<td>73.6</td>
<td><strong>81.4</strong></td>
<td><strong>81.4</strong></td>
<td><strong>77.9</strong></td>
<td><strong>77.9</strong></td>
<td><strong>79.6</strong></td>
<td><strong>79.6</strong></td>
<td><strong>79.3</strong></td>
<td><strong>79.3</strong></td>
<td><strong>75.7</strong></td>
<td><strong>75.7</strong></td>
</tr>
<tr>
<td>HT-D (NN) J=0</td>
<td>87.9</td>
<td>87.9</td>
<td>82.1</td>
<td>82.1</td>
<td>75.7</td>
<td>75.7</td>
<td>81.1</td>
<td>81.1</td>
<td>76.8</td>
<td>76.8</td>
<td>76.1</td>
<td>76.1</td>
<td>81.1</td>
<td>81.1</td>
<td>75.0</td>
<td>75.0</td>
</tr>
<tr>
<td>HT-D (NN) J=1</td>
<td>88.6</td>
<td>88.6</td>
<td>84.6</td>
<td>84.6</td>
<td>76.8</td>
<td>76.8</td>
<td>79.6</td>
<td>79.6</td>
<td>75.4</td>
<td>75.4</td>
<td>78.6</td>
<td>78.6</td>
<td>82.9</td>
<td>82.9</td>
<td>76.1</td>
<td>76.1</td>
</tr>
<tr>
<td>HT-D (NN) J=2</td>
<td>90.0</td>
<td>90.0</td>
<td>82.1</td>
<td>82.1</td>
<td>73.9</td>
<td>73.9</td>
<td>80.7</td>
<td>80.7</td>
<td>81.1</td>
<td>81.1</td>
<td>81.8</td>
<td>81.8</td>
<td>83.9</td>
<td>83.9</td>
<td>74.6</td>
<td>74.6</td>
</tr>
</tbody>
</table>
Results with Neural Features

Data: Multilingual Geoquery

Baselines:  (Lu et al., 2008) (Lu, 2015) (Susanto and Lu, 2017)

(+o): models with distributed representations of semantic units.

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
<th>F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT-G</td>
<td>76.8</td>
<td>81.0</td>
</tr>
<tr>
<td>HT-D</td>
<td>86.8</td>
<td>86.8</td>
</tr>
<tr>
<td>HT-D (+O)</td>
<td>86.1</td>
<td>86.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>English</th>
<th>Thai</th>
<th>German</th>
<th>Greek</th>
<th>Chinese</th>
<th>Indonesian</th>
<th>Swedish</th>
<th>Farsi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
<td>Acc.</td>
<td>F.</td>
</tr>
<tr>
<td>HT-G</td>
<td>76.8</td>
<td>81.0</td>
<td>73.6</td>
<td>76.7</td>
<td>62.1</td>
<td>68.5</td>
<td>69.3</td>
<td>74.6</td>
</tr>
<tr>
<td>HT-D</td>
<td>86.8</td>
<td>86.8</td>
<td>80.7</td>
<td>80.7</td>
<td>75.7</td>
<td>75.7</td>
<td>79.3</td>
<td>79.3</td>
</tr>
<tr>
<td>HT-D (+O)</td>
<td>86.1</td>
<td>86.1</td>
<td>81.1</td>
<td>81.1</td>
<td>73.6</td>
<td>73.6</td>
<td>81.4</td>
<td>81.4</td>
</tr>
<tr>
<td>HT-D (NN) J=0</td>
<td>87.9</td>
<td>87.9</td>
<td>82.1</td>
<td>82.1</td>
<td>75.7</td>
<td>75.7</td>
<td>81.1</td>
<td>81.1</td>
</tr>
<tr>
<td>HT-D (NN) J=1</td>
<td>88.6</td>
<td>88.6</td>
<td>84.6</td>
<td>84.6</td>
<td>76.8</td>
<td>76.8</td>
<td>79.6</td>
<td>79.6</td>
</tr>
<tr>
<td>HT-D (NN) J=2</td>
<td>90.0</td>
<td>90.0</td>
<td>82.1</td>
<td>82.1</td>
<td>73.9</td>
<td>73.9</td>
<td>80.7</td>
<td>80.7</td>
</tr>
<tr>
<td>HT-D (NN+O) J=0</td>
<td>86.1</td>
<td>86.1</td>
<td>83.6</td>
<td>83.6</td>
<td>73.9</td>
<td>73.9</td>
<td>82.1</td>
<td>82.1</td>
</tr>
<tr>
<td>HT-D (NN+O) J=1</td>
<td>86.1</td>
<td>86.1</td>
<td>86.1</td>
<td>86.1</td>
<td>72.5</td>
<td>72.5</td>
<td>80.4</td>
<td>80.4</td>
</tr>
<tr>
<td>HT-D (NN+O) J=2</td>
<td>89.6</td>
<td>86.1</td>
<td>84.6</td>
<td>84.6</td>
<td>72.1</td>
<td>72.1</td>
<td>83.2</td>
<td>83.2</td>
</tr>
</tbody>
</table>
## Results with Neural Features

**Data:** Multilingual Geoquery  
**Baselines:** (Lu et al., 2008) (Lu, 2015) (Susanto and Lu, 2017)  
(+o): models with distributed representations of semantic units.

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
<th>F.</th>
<th>Acc.</th>
<th>F.</th>
<th>Acc.</th>
<th>F.</th>
<th>Acc.</th>
<th>F.</th>
<th>Acc.</th>
<th>F.</th>
<th>Acc.</th>
<th>F.</th>
<th>Acc.</th>
<th>F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT-G</td>
<td>76.8</td>
<td>81.0</td>
<td>73.6</td>
<td>76.7</td>
<td>62.1</td>
<td>68.5</td>
<td>69.3</td>
<td>74.6</td>
<td>56.1</td>
<td>58.4</td>
<td>66.4</td>
<td>72.8</td>
<td>61.4</td>
<td>70.5</td>
</tr>
<tr>
<td>HT-D (+O)</td>
<td>86.8</td>
<td>86.8</td>
<td>80.7</td>
<td>80.7</td>
<td>75.7</td>
<td>75.7</td>
<td>79.3</td>
<td>79.3</td>
<td>76.1</td>
<td>76.1</td>
<td>75.0</td>
<td>75.0</td>
<td><strong>79.3</strong></td>
<td><strong>79.3</strong></td>
</tr>
<tr>
<td>HT-D (NN) J=0</td>
<td>86.1</td>
<td>86.1</td>
<td><strong>81.1</strong></td>
<td><strong>81.1</strong></td>
<td>73.6</td>
<td>73.6</td>
<td><strong>81.4</strong></td>
<td><strong>81.4</strong></td>
<td>77.9</td>
<td>77.9</td>
<td><strong>79.6</strong></td>
<td><strong>79.6</strong></td>
<td>79.3</td>
<td>79.3</td>
</tr>
<tr>
<td>HT-D (NN) J=1</td>
<td>87.9</td>
<td>87.9</td>
<td>82.1</td>
<td>82.1</td>
<td>75.7</td>
<td>75.7</td>
<td>81.1</td>
<td>81.1</td>
<td>76.8</td>
<td>76.8</td>
<td>76.1</td>
<td>76.1</td>
<td>81.1</td>
<td>81.1</td>
</tr>
<tr>
<td>HT-D (NN) J=2</td>
<td>90.0</td>
<td>90.0</td>
<td>82.1</td>
<td>82.1</td>
<td>73.9</td>
<td>73.9</td>
<td>80.7</td>
<td>80.7</td>
<td>81.1</td>
<td>81.1</td>
<td>81.8</td>
<td>81.8</td>
<td><strong>83.9</strong></td>
<td><strong>83.9</strong></td>
</tr>
<tr>
<td>HT-D (NN+O) J=0</td>
<td>86.1</td>
<td>86.1</td>
<td>83.6</td>
<td>83.6</td>
<td>73.9</td>
<td>73.9</td>
<td>82.1</td>
<td>82.1</td>
<td>77.9</td>
<td>77.9</td>
<td>81.1</td>
<td>81.1</td>
<td>82.1</td>
<td>82.1</td>
</tr>
<tr>
<td>HT-D (NN+O) J=1</td>
<td>86.1</td>
<td>86.1</td>
<td><strong>86.1</strong></td>
<td><strong>86.1</strong></td>
<td>72.5</td>
<td>72.5</td>
<td>80.4</td>
<td>80.4</td>
<td>81.4</td>
<td>81.4</td>
<td>82.5</td>
<td>82.5</td>
<td>82.5</td>
<td>82.5</td>
</tr>
<tr>
<td>HT-D (NN+O) J=2</td>
<td>89.6</td>
<td>86.1</td>
<td>84.6</td>
<td>84.6</td>
<td>72.1</td>
<td>72.1</td>
<td><strong>83.2</strong></td>
<td><strong>83.2</strong></td>
<td><strong>82.1</strong></td>
<td><strong>82.1</strong></td>
<td><strong>83.9</strong></td>
<td><strong>83.9</strong></td>
<td>83.6</td>
<td>83.6</td>
</tr>
</tbody>
</table>

5 out of 8 languages get improved
Semantic units with similar meanings gather together.

Occasionally, semantic units conveying opposite meanings are grouped together.
Conclusions

✓ Summary
  ✓ Presented a novel method to learning distributed representations of semantic units containing cross-lingual information.
Conclusions

✓ Summary
  ✓ Presented a novel method to learning distributed representations of semantic units containing cross-lingual information.

✓ Future work
  ✓ Learn representations and semantic parsers in a joint manner.
Conclusions

✓ Summary
  ✓ Presented a novel method to learning distributed representations of semantic units containing cross-lingual information.

✓ Future work
  ✓ Learn representations and semantic parsers in a joint manner.
  ✓ Investigate which languages from auxiliary corpus are the leading sources of performance gains.
Code available at: http://statnlp.org/research/sp/

Questions?