Driving Semantic Parsing from the World’s Response

James Clarke, Dan Goldwasser, Ming-Wei Chang, Dan Roth

Cognitive Computation Group
University of Illinois at Urbana-Champaign

CoNLL 2010
What is Semantic Parsing?

Meaning Representation

I'd like a coffee with no sugar and just a little milk

\texttt{make(coffee, sugar=0, milk=0.3)}
What is Semantic Parsing?

Meaning Representation

I’d like a coffee with no sugar and just a little milk

make(coffee, sugar=0, milk=0.3)
Supervised Learning Problem

Challenges:

- Structured Prediction problem
- Model part of the structure as hidden?
Multiple approaches to the problem:

- **KRISP** (Kate & Mooney 2006)
  - SVM-based parser using string kernels.
- **Zettlemoyer & Collins 2005; Zettlemoyer & Collins 2007**
  - Probabilistic parser based on relaxed CCG grammars.
- **WASP** (Wong & Mooney 2006; Wong & Mooney 2007)
  - Based on Synchronous CFG.
- **Ge & Mooney 2009**
  - Integrated syntactic and semantic parser.
Multiple approaches to the problem:

- **KRISP (Kate & Mooney 2006)**
  - SVM-based parser using string kernels.
- **Zettlemoyer & Collins 2005; Zettlemoyer & Collins 2007**
  - Probabilistic parser based on relaxed CCG grammars.
- **WASP (Wong & Mooney 2006; Wong & Mooney 2007)**
  - Based on Synchronous CFG.
- **Ge & Mooney 2009**
  - Integrated syntactic and semantic parser.

**Assumption**: A training set consisting of natural language and meaning representation pairs.
Using the World’s response

Meaning Representation

make(coffee, sugar=0, milk=0.3)

I’d like a coffee with no sugar and just a little milk
Using the World’s response

Meaning Representation

make(coffee, sugar=0, milk=0.3)

I’d like a coffee with no sugar and just a little milk

Good!

Bad!
Using the World’s response

Meaning Representation

I’d like a coffee with no sugar and just a little milk

make(coffee, sugar=0, milk=0.3)

Good!

Bad!

Question: Can we use feedback based on the response to provide supervision?
We aim to:
- Reduce the burden of annotation for semantic parsing.

We focus on:
- Using the World’s response to learn a semantic parser.
- Developing new training algorithms to support this learning paradigm.
- A lightweight semantic parsing model that doesn’t require annotated data.

This results in:
- Learning a semantic parser using zero annotated meaning representations.
Outline

1. Semantic Parsing

2. Learning
   - DIRECT Approach
   - AGGRESSIVE Approach

3. Semantic Parsing Model

4. Experiments
Outline

1 Semantic Parsing

2 Learning
   - DIRECT Approach
   - AGGRESSIVE Approach

3 Semantic Parsing Model

4 Experiments
What is the largest state that borders Texas?

\[ \text{largest(state(next_to(texas)))} \]
What is the largest state that borders Texas?

\[
\text{largest}(\text{state(next_to(texas))})
\]

\[
\hat{z} = F_w(x) = \arg \max_{y \in \mathcal{Y}, z \in \mathcal{Z}} w^T \phi(x, y, z)
\]

\[F : \mathcal{X} \rightarrow \mathcal{Z}\]
What is the largest state that borders Texas?

\[ \text{largest}(\text{state}(\text{next_to}(\text{texas}))) \]

\[ F : \mathcal{X} \rightarrow \mathcal{Z} \]

\[ \hat{z} = F_w(x) = \arg \max_{y \in \mathcal{Y}, z \in \mathcal{Z}} w^T \phi(x, y, z) \]

- **Model** The nature of inference and feature functions.
- **Learning Strategy** How we obtain the weights.
What is the largest state that borders Texas?

\[
\text{largest(\text{state(next\_to(texas))})}
\]

Response \( r \)

New Mexico

\[
F : \mathcal{X} \rightarrow \mathcal{Z}
\]

\[
\hat{z} = F_w(x) = \arg\max_{y \in \mathcal{Y}, z \in \mathcal{Z}} w^T \Phi(x, y, z)
\]

- **Model** The nature of inference and feature functions.
- **Learning Strategy** How we obtain the weights.
Outline

1. Semantic Parsing

2. Learning
   - DIRECT Approach
   - AGGRESSIVE Approach

3. Semantic Parsing Model

4. Experiments
Learning

Inputs:

- Natural language sentences.
- \( \text{Feedback} : \mathcal{X} \times \mathcal{Z} \rightarrow \{+1, -1\} \).
- Zero meaning representations.
Learning

Inputs:

- Natural language sentences.
- Feedback: $\mathcal{X} \times \mathcal{Z} \rightarrow \{+1, -1\}$.
- Zero meaning representations.

\[
\text{Feedback}(x, z) = \begin{cases} 
+1 & \text{if } \text{execute}(z) = r \\
-1 & \text{otherwise}
\end{cases}
\]
Learning

Inputs:
- Natural language sentences.
- Feedback: $\mathcal{X} \times \mathcal{Z} \rightarrow \{+1, -1\}$.
- Zero meaning representations.

Goal: A weight vector that scores the correct meaning representation higher than all other meaning representations.

Response Driven Learning:

Input text \arrow{predict} Meaning Representation \arrow{apply to} World
Learning Strategies

\[ \text{repeat} \quad \text{for all input sentences do} \]
\[ \quad \text{Solve the inference problem} \]
\[ \quad \text{Query } Feedback \text{ function} \]
\[ \text{end for} \]
\[ \text{Learn a new } w \text{ using feedback} \]
\[ \text{until Convergence} \]
repeat
for all input sentences do
    Solve the inference problem
    Query Feedback function
end for
Learn a new \( w \) using feedback
until Convergence

\[ y, z = \arg \max w^T \Phi(x, y, z) \]
repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
until Convergence

Learn a new \( \mathbf{w} \) using feedback

\[
\begin{align*}
\mathbf{x}_1 &\rightarrow \mathbf{y}_1 \rightarrow \mathbf{z}_1 \rightarrow +1 \\
\mathbf{x}_2 &\rightarrow \mathbf{y}_2 \rightarrow \mathbf{z}_2 \rightarrow -1 \\
\mathbf{x}_3 &\rightarrow \mathbf{y}_3 \rightarrow \mathbf{z}_3 \rightarrow -1 \\
\vdots &\rightarrow \vdots \rightarrow \vdots \\
\mathbf{x}_n &\rightarrow \mathbf{y}_n \rightarrow \mathbf{z}_n \rightarrow -1
\end{align*}
\]
Learning Strategies

repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
  Learn a new $w$ using feedback
until Convergence

$y = \arg \max_w w^T \Phi(x, y, z)$

Clarke, Goldwasser, Chang, Roth
Outline

1. Semantic Parsing

2. Learning
   - DIRECT Approach
   - AGGRESSIVE Approach

3. Semantic Parsing Model

4. Experiments
DIRECT Approach

Learn a binary classifier to discriminate between good and bad meaning representations.
DIRECT Approach

Use \((x, y, z)\) as a training example with label from feedback.
**DIRECT Approach**

- Use \((x, y, z)\) as a training example with label from feedback.
- Find \(w\) such that \(f \cdot w^T \Phi(x, y, z) > 0\)
DIRECT Approach

Each point represented by $\Phi(x, y, x)$ normalized by $|x|$
Learn a binary classifier to discriminate between good and bad meaning representations.
DIRECT Approach

repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
  Learn a new \( w \) using feedback
until Convergence
DIRECT Approach

repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
  Learn a new $w$ using feedback
until Convergence

$\mathbf{x}_1$

$\mathbf{x}_2$

$\mathbf{x}_3$

\vdots

$\mathbf{x}_n$
**DIRECT Approach**

```
repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
  Learn a new $w$ using feedback
until Convergence

$y, z = \arg \max w^T \Phi(x, y, z)$
```
DIRECT Approach

\[
x_1 \rightarrow y'_1 \rightarrow z'_1 \rightarrow +1
\]

\[
x_2 \rightarrow y'_2 \rightarrow z'_2 \rightarrow +1
\]

\[
x_3 \rightarrow y'_3 \rightarrow z'_3 \rightarrow +1
\]

\[
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots
\]

\[
x_n \rightarrow y'_n \rightarrow z'_n \rightarrow -1
\]

repeat

for all input sentences do

Solve the inference problem

Query Feedback function

end for

Learn a new \( w \) using feedback

until Convergence
DIRECT Approach

\[ \begin{align*}
  x_1, y'_1, z'_1 & \rightarrow +1 \\
  x_2, y'_2, z'_2 & \rightarrow +1 \\
  x_3, y'_3, z'_3 & \rightarrow +1 \\
  \vdots & \quad \vdots \\
  x_n, y'_n, z'_n & \rightarrow -1
\end{align*} \]

repeat

for all input sentences do
  Solve the inference problem
  Query Feedback function
end for

Learn a new \( w \) using feedback

until Convergence

Clarke, Goldwasser, Chang, Roth
DIRECT Approach

Repeat until convergence!
DIRECT Approach

Repeat until convergence!
DIRECT Approach

Repeat until convergence!
DIRECT Approach

Repeat until convergence!
Outline

1. Semantic Parsing

2. Learning
   - DIRECT Approach
   - AGGRESSIVE Approach

3. Semantic Parsing Model

4. Experiments
AGGRESSIVE Approach

Positive feedback is a good indicator of the correct meaning representation.

Use data with positive feedback as training data for structured learning.
AGGRESSIVE Approach

\[
\begin{align*}
  x_1 &\rightarrow y_1 \rightarrow z_1 \rightarrow +1 \\
  x_2 &\rightarrow y_2 \rightarrow z_2 \rightarrow -1 \\
  x_3 &\rightarrow y_3 \rightarrow z_3 \rightarrow -1 \\
  \vdots & \vdots \vdots \vdots \quad \vdots \\
  x_n &\rightarrow y_n \rightarrow z_n \rightarrow -1
\end{align*}
\]

repeat
for all input sentences do
  Solve the inference problem
  Query \textit{Feedback} function
end for
Learn a new \textbf{w} using feedback
until Convergence
AGGRESSIVE Approach

- Use items with positive feedback as training data for a structured learner.

\[
\begin{align*}
\mathbf{x}_1 & \rightarrow \mathbf{y}_1 \rightarrow \mathbf{z}_1 \rightarrow +1 \\
\mathbf{x}_2 & \rightarrow \mathbf{y}_2 \rightarrow \mathbf{z}_2 \rightarrow -1 \\
\mathbf{x}_3 & \rightarrow \mathbf{y}_3 \rightarrow \mathbf{z}_3 \rightarrow -1 \\
\vdots & \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
\mathbf{x}_n & \rightarrow \mathbf{y}_n \rightarrow \mathbf{z}_n \rightarrow -1
\end{align*}
\]
AGGRESSIVE Approach

Use items with positive feedback as training data for a structured learner.
AGGRESSIVE Approach

Use items with positive feedback as training data for a structured learner.

\[ \mathbf{x}_1 \rightarrow \mathbf{y}_1 \rightarrow \mathbf{z}_1 \]

\[ \mathbf{x}_2 \rightarrow \mathbf{y}_2 \rightarrow \mathbf{z}_2 \]

\[ \mathbf{x}_3 \rightarrow \mathbf{y}_3 \rightarrow \mathbf{z}_3 \]

\[ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \]

\[ \mathbf{x}_n \rightarrow \mathbf{y}_n \rightarrow \mathbf{z}_n \]
AGGRESSIVE Approach

Use items with positive feedback as training data for a structured learner.

Implicitly consider all other meaning representations for these examples as bad.

Find $\mathbf{w}$ such that

$$\mathbf{w}^T \Phi(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*) > \mathbf{w}^T \Phi(\mathbf{x}, \mathbf{y}', \mathbf{z}')$$
AGGRESSIVE Approach

repeat
for all input sentences do
  Solve the inference problem
  Query Feedback function
end for
Learn a new \( w \) using feedback
until Convergence
AGGRESSIVE Approach

\[
\begin{align*}
x_1, x_2, x_3, \ldots, x_n
\end{align*}
\]

repeat
for all input sentences do
  Solve the inference problem
  Query Feedback function
end for
Learn a new \( w \) using feedback
until Convergence

Clarke, Goldwasser, Chang, Roth
AGGRESSIVE Approach

\[ x_1 \rightarrow y'_1 \rightarrow z'_1 \]

\[ x_2 \rightarrow y'_2 \rightarrow z'_2 \]

\[ x_3 \rightarrow y'_3 \rightarrow z'_3 \]

\[ \vdots \]

\[ x_n \rightarrow y'_n \rightarrow z'_n \]

\[ \text{repeat} \]
\[ \text{for all input sentences do} \]
\[ \text{Solve the inference problem} \]
\[ \text{Query Feedback function} \]
\[ \text{end for} \]
\[ \text{Learn a new } w \text{ using feedback} \]
\[ \text{until Convergence} \]

\[ y, z = \arg \max w^T \Phi(x, y, z) \]
repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
  Learn a new \( w \) using feedback
until Convergence
**AGGRESSIVE Approach**

repeat
  for all input sentences do
    Solve the inference problem
    Query *Feedback* function
  end for
  Learn a new $w$ using feedback
until Convergence
AGGRESSIVE Approach

repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
  Learn a new \( \mathbf{w} \) using feedback
until Convergence

\[ \mathbf{x}_1 \rightarrow \mathbf{y}_1' \rightarrow \mathbf{z}_1' \]
\[ \mathbf{x}_2 \rightarrow \mathbf{y}_2' \rightarrow \mathbf{z}_2' \]
\[ \mathbf{x}_3 \rightarrow \mathbf{y}_3' \rightarrow \mathbf{z}_3' \]
\[ \vdots \rightarrow \vdots \rightarrow \vdots \]
\[ \mathbf{x}_n \rightarrow \mathbf{y}_n' \rightarrow \mathbf{z}_n' \]
AGGRESSIVE Approach

\[
\begin{align*}
  x_1 &\quad \rightarrow \quad y'_1 \quad z'_1 \\
  x_2 &\quad \rightarrow \quad y'_2 \quad z'_2 \\
  x_3 &\quad \rightarrow \quad y'_3 \quad z'_3 \\
  \vdots &\quad \quad \vdots \quad \quad \vdots \\
  x_n &\quad \rightarrow \quad y'_n \quad z'_n
\end{align*}
\]

repeat
  for all input sentences do
    Solve the inference problem
    Query Feedback function
  end for
  Learn a new \( w \) using feedback
until Convergence

Clark, Goldwasser, Chang, Roth
Summary of Learning Strategies

- **DIRECT** Uses both positive and negative feedback as examples to train a binary classifier.
- **AGGRESSIVE** Adapts the feedback signal and uses only positive feedback to train a structured predictor.

Clarke, Goldwasser, Chang, Roth
Outline

1. Semantic Parsing

2. Learning
   - DIRECT Approach
   - AGGRESSIVE Approach

3. Semantic Parsing Model

4. Experiments
What is the largest state that borders Texas?

\[
\hat{z} = F_w(x) = \arg \max_{y \in Y, z \in Z} w^T \Phi(x, y, z)
\]

- **First-order**: Map lexical items. \textit{largest} $\rightarrow$ \textit{largest}
- **Second-order**: Composition. \textit{next\_to} (\textit{state} (·)) or \textit{state} (\textit{next\_to} (·))

Inference procedure leverages the typing information of the domain.
How many people live in the state of Texas?

**Goal:** \( \text{population}(\text{state(texas)}) \)
How many people live in the state of Texas?

Goal: `population(state(texas))`
First-order Decisions

How many people live in the state of Texas?

Goal: population(state(texas))
First-order Decisions

How many people live in the state of Texas?

- Use a simple lexicon to bootstrap the process.

Goal: \text{population(state(texas))}
How many people live in the state of Texas?

- Use a simple lexicon to bootstrap the process.

**Goal:** population(state(texas))
First-order Decisions

How many people live in the state of Texas?

- Use a simple lexicon to bootstrap the process.

Goal: \( \text{population}(\text{state}('texas')) \)
First-order Decisions

How many people live in the state of Texas?

- Use a simple lexicon to bootstrap the process.
- Lexical resources help us move beyond the lexicon.

\[ \text{wordnet_sim(people,population)} \]

**Goal:** \( \text{population(state(texas))} \)
Use a simple lexicon to bootstrap the process.

Lexical resources help us move beyond the lexicon.

```python
wordnet_sim(people, population)
```

Goal: population(state(texas))
How many people live in the state of Texas?

- Use a simple lexicon to bootstrap the process.
- Lexical resources help us move beyond the lexicon.
  \[
  \text{wordnet\_sim}(\text{people}, \text{population})
  \]
- Context helps disambiguate between choices.

\textbf{Goal:} \text{population}(\text{state(texas)})
Second-order Decisions

How do we compose the predicates and constants.

**Domain dependent:**
- Encode typing information inherent in the domain into the inference procedure.
- \( \text{population(state(\cdot)) vs state(population(\cdot))} \)

**Features:**
- Dependency path distance.
- Word position distance.
- Predicate “bigrams”.
- \( \text{next_to(state(\cdot)) vs state(next_to(\cdot))} \)
Outline

1. Semantic Parsing

2. Learning
   - DIRECT Approach
   - AGGRESSIVE Approach

3. Semantic Parsing Model

4. Experiments
Domain:
GEOQUERY U.S Geographical Questions.
- Response 250. \((x, r)\) pairs. Zero meaning representations.
- Query 250. \((x)\) sentences.

Evaluation metric:
Accuracy (percentage of meaning representations that return the correct answer).
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>R250</th>
<th>Q250</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOLEARN</td>
<td>22.2</td>
<td>—</td>
</tr>
<tr>
<td>DIRECT AGGRESSIVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUPERVISED</td>
<td>87.6</td>
<td>80.4</td>
</tr>
<tr>
<td>Algorithm</td>
<td>R250</td>
<td>Q250</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>NOLEARN</td>
<td>22.2</td>
<td>—</td>
</tr>
<tr>
<td>DIRECT AGGRESSIVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUPERVISED</td>
<td>87.6</td>
<td>80.4</td>
</tr>
</tbody>
</table>

- NOLEARN used to initialize both learning approaches.
Q: How good is our model when trained in a fully supervised manner?

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>R250</th>
<th>Q250</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOLEARN</td>
<td>22.2</td>
<td>—</td>
</tr>
<tr>
<td>DIRECT AGGRESSIVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUPERVISED</td>
<td>87.6</td>
<td>80.4</td>
</tr>
</tbody>
</table>
Q: How good is our model when trained in a fully supervised manner?

A: 80% on test data. Other supervised methods range from 60% to 85% accuracy.
## Learning Behavior

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>R250</th>
<th>Q250</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NO LEARN</strong></td>
<td>22.2</td>
<td>—</td>
</tr>
<tr>
<td><strong>DIRECT AGGRESSIVE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SUPERVISED</strong></td>
<td>87.6</td>
<td>80.4</td>
</tr>
</tbody>
</table>

**Q:** Is it possible to learn without any meaning representations?
## Learning Behavior

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>R250</th>
<th>Q250</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NO LEARN</strong></td>
<td>22.2</td>
<td></td>
</tr>
<tr>
<td><strong>DIRECT</strong></td>
<td>75.2</td>
<td>69.2</td>
</tr>
<tr>
<td><strong>AGGRESSIVE</strong></td>
<td>82.4</td>
<td>73.2</td>
</tr>
<tr>
<td><strong>SUPERVISED</strong></td>
<td>87.6</td>
<td>80.4</td>
</tr>
</tbody>
</table>

- **Q:** Is it possible to learn without any meaning representations?  
- **A:** Yes!  
- **A:** Learns to cover more of the Response data set.  
- **A:** And only 7% below the SUPERVISED upper bound.
Learning Behavior

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>R250</th>
<th>Q250</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOLEARN</td>
<td>22.2</td>
<td>—</td>
</tr>
<tr>
<td>DIRECT</td>
<td>75.2</td>
<td>69.2</td>
</tr>
<tr>
<td>AGGRESSIVE</td>
<td>82.4</td>
<td>73.2</td>
</tr>
<tr>
<td>SUPERVISED</td>
<td>87.6</td>
<td>80.4</td>
</tr>
</tbody>
</table>

**Q:** Is it possible to learn without any meaning representations?

**A:** Yes!

**A:** Learns to cover more of the Response data set.

**A:** And only 7% below the SUPERVISED upper bound.
AGGRESSIVE correctly interprets 16% that DIRECT does not. 9% vice-versa. Leaving only 9% incorrect.
Similar to indirect learning protocols:

- Learning a binary classifier with “hidden explanation”. Supervision only required for binary data. No labeled structures. NAACL 2010 (Chang, Goldwasser, Roth, Srikumar 2010a).

- Structured learning with binary and structured labels. Mix of supervision for binary data and structured data. Binary label indicates whether input has a “good” structure. ICML 2010 (Chang, Goldwasser, Roth, Srikumar 2010b).
Conclusions

Contributions:

- Response Driven Learning. A new learning paradigm that doesn’t rely on annotated meaning representations. Supervised at the response level. Natural supervision signal.
- Two learning algorithms capable of working within response driven learning.
- A shallow semantic parsing model.

Future work:

- Can we combine the two learning algorithms?
- Other semantic parsing domains?
- Response driven learning for other tasks?