Alignment-Based Compositional Semantics for Instruction Following

Jacob Andreas and Dan Klein
UC Berkeley
Following instructions

go down the green hall

go all the way

then face the chair
Following instructions

go down the green hall

go all the way

then face the chair
Following instructions: parsing

go down the green hall

go all the way

then face the chair
Following instructions: parsing

- go down the green hall
- go all the way
- then *face the chair*
Following instructions: planning

- go down the green hall
- go all the way
- then face the chair
Following instructions: planning

*go down the green hall*

*go all the way*

*then face the chair*
Following instructions

Planning
(no instruction)
[ go down the hall, go all the way ]

Parsing

green hall
face the chair
Previous work

Planning
Branavan et al. (2009)
Vogel and Jurafsky (2010)

Parsing
Chen and Mooney (2011)
Artzi and Zettlemoyer (2013)
Kim and Mooney (2013)
Tellex et al. (variously)
Parsing and planning

Sequence-to-sequence

go down the green hall  turn left

Structure-to-structure

go down the green hall
SEQUENCE-TO-SEQUENCE MODEL
go down the yellow hall  turn left
Sequence-to-sequence

go down the yellow hall  turn left

Alignments
go down the yellow hall  turn left

Alignments

Actions
go down the yellow hall  turn left

Sequence-to-sequence

Alignments

Actions
go down the yellow hall  turn left

Sequence-to-sequence

Alignments

Actions
go down the yellow hall  turn left

Sequence-to-sequence

Alignments

Actions
Sequence-to-sequence

go down the yellow hall  turn left
Sequence-to-sequence

go down the yellow hall  turn left

1  2  3
go down the yellow hall  turn left

Sequence-to-sequence
Sequence-to-sequence

*go down the yellow hall*

*turn left*
go down the yellow hall  turn left
Sequence-to-sequence

go down the yellow hall  turn left
STRUCTURE-TO-STRUCTURE MODEL
go down the yellow hall
Structure-to-structure

go → down → hall

yellow

the
go → down → hall → yellow → the
Structure-to-structure

go down

the yellow hall
go down the yellow hall
Structure-to-structure

down

good

down

tal

yellow
JOINT MODEL
Joint model

go down the yellow hall  turn left
go down the yellow hall  turn left
Joint model

go down the yellow hall  turn left
Joint model

go down the yellow hall

turn left
Joint model

\[ \psi \left( \begin{array}{c}
\text{go down} \\
\text{the yellow hall}
\end{array} \right) = \theta^\top \left( \phi(\text{go}, \cdot) + \phi(\text{down}, \cdot) + \cdots \right) \]
Joint model

\[ p(\text{actions, alignment} \mid \text{text}; \theta) \propto \exp \sum \psi \left( \begin{array}{c}
\text{go down} \\
\text{the yellow hall}
\end{array} \right) \]
go down the yellow hall

turn left
LEARNING / INFERENCE
Learning: coordinate ascent

$$\max_{\theta, \text{alignments}} p(\text{actions, alignments | text; } \theta)$$

$$\max \text{ alignments using custom alignment dynamic program}$$

$$\max_{\theta} \text{ using L-BFGS}$$
Decoding: coordinate ascent

\[ p(\text{actions}, \text{alignments} \mid \text{text}; \theta) \]

\[
\text{max} \\
\text{actions, alignments}
\]

\[
\text{max} \\
\text{alignments}
\]

using custom alignment dynamic program

\[
\text{max} \\
\text{actions}
\]

using a planner
EXPERIMENTS
Three tasks
Puzzle solving

*clear the two long columns, and then the row*

[Branavan+ 09]
Puzzle solving

clear the two long columns, and then the row

[Branavan+ 09]
Puzzle solving

Plan execution (%)
Puzzle solving

Task completion (%)

<table>
<thead>
<tr>
<th></th>
<th>Task completion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No text</td>
<td>78</td>
</tr>
<tr>
<td>This work</td>
<td>86</td>
</tr>
</tbody>
</table>
Maze navigation

you should see a grey floor to your right and a chair in front of you

[MacMahon+ 91]
Maze navigation

Task completion (%)
right round the white water [...] but stay quite close ’cause you don’t otherwise you’re going to be in that stone creek

[Anderson+ 91]
Map reading

Plan execution ($F_0$)

<table>
<thead>
<tr>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
Ablations: maze navigation

Task completion (%)

Without grounding graphs

This work

60
Ablations: map reading

Plan execution ($F_0$)

- 60

This work
Conclusion

Structured alignment/decoding gives us best aspects of:

- Compositional semantics (like a parser)
- Sequence structure (like a planner)
THANK YOU