Modeling Discourse Cohesion for Discourse Parsing via Memory Network

Yanyan Jia, Yuan Ye, Yansong Feng, Yuxuan Lai, Rui Yan and Dongyan Zhao

Institute of Computer Science and Technology, Peking University
Discourse Dependency Parsing

EDU_1: President Bush insists
EDU_2: it would be a great tool
EDU_3: for curbing the budget deficit
EDU_4: and slicing the lard out of government programs.
EDU_5: He wants it now.

... 

EDU_32: Mr. Bush is considering simply declaring
EDU_33: that the Constitution gives him the power

...
Discourse Dependency Parsing

**EDU_1**: President Bush insists

**EDU_2**: it would be a great tool

**EDU_3**: for curbing the budget deficit

**EDU_4**: and slicing the lard out of government programs.

**EDU_5**: He wants it now.

…

**EDU_32**: Mr. Bush is considering simply declaring

**EDU_33**: that the Constitution gives him the power

…
Motivation

• Identifying **long-span dependencies** between element discourse units
  
  – **Discourse structure**
    • Morris and Hirst, 1991 extracts features to characterize discourse structures
  
  – **Discourse cohesion**
    • Joty et al., 2013 uses lexical chain features to model discourse cohesion
Motivation

• Identifying long-span dependencies between element discourse units

  – Discourse structure
    • Morris and Hirst, 1991 extracts features to characterize discourse structures
  – Discourse cohesion
    • Joty et al., 2013 uses lexical chain feature to model discourse cohesion

Our Work: Use Memory network to implicitly capture discourse cohesion
How Does Memory Network Work?

**EDU$_1$:** I feel hungry after wake up,
**EDU$_2$:** I rush into the kitchen and make my breakfast.
**EDU$_3$:** My breakfast is hamburger.

**EDU$_4$:** It is eight o’clock when I leave home.
**EDU$_5$:** So late!

**EDU$_6$:** I drive into the highway,
**EDU$_7$:** but meet a traffic jam.
**EDU$_8$:** Oh, I finally arrive at the company.

**EDU$_9$:** It is nine o’clock.
**EDU$_{10}$:** Thank God, I am not late for work.

**EDU$_{11}$:** But the hamburger is cold,
**EDU$_{12}$:** order some take-away food is better, maybe.
How Does Memory Network Work?

**EDU_1**: I feel hungry after wake up,
**EDU_2**: I rush into the kitchen and make my breakfast.
**EDU_3**: My breakfast is hamburger.

**EDU_4**: It is eight o’clock when I leave home.
**EDU_5**: So late!

**EDU_6**: I drive into the highway,
**EDU_7**: but meet a traffic jam.
**EDU_8**: Oh, I finally arrive at the company.

**EDU_9**: It is nine o’clock.
**EDU_10**: Thank God, I am not late for work.

**EDU_11**: But the hamburger is cold,
**EDU_12**: order some take-away food is better, maybe.
How Does Memory Network Work?

EDU₁: I feel hungry after wake up,
EDU₂: I rush into the kitchen and make my breakfast.
EDU₃: My breakfast is hamburger.

EDU₄: It is eight o’clock when I leave home.
EDU₅: So late!

EDU₆: I drive into the highway,
EDU₇: but meet a traffic jam.
EDU₈: Oh, I finally arrive at the company.

EDU₉: It is nine o’clock.
EDU₁₀: Thank God, I am not late for work.

EDU₁₁: But the hamburger is cold,
EDU₁₂: order some take-away food is better, maybe.
How Does Memory Network Work?

EDU_1: I feel hungry after wake up,
EDU_2: I rush into the kitchen and make my breakfast.
EDU_3: My breakfast is hamburger.

EDU_4: It is eight o’clock when I leave home.
EDU_5: So late!

EDU_6: I drive into the highway,
EDU_7: but meet a traffic jam.
EDU_8: Oh, I finally arrive at the company.

EDU_9: It is nine o’clock.
EDU_10: Thank God, I am not late for work.

EDU_11: But the hamburger is cold,
EDU_12: order some take-away food is better, maybe.
How Does Memory Network Work?

**EDU_1**: I feel hungry after wake up.
**EDU_2**: I rush into the kitchen and make my breakfast.
**EDU_3**: My breakfast is hamburger.

**EDU_4**: It is eight o’clock when I leave home.
**EDU_5**: So late!

**EDU_6**: I drive into the highway,
**EDU_7**: but meet a traffic jam.
**EDU_8**: Oh, I finally arrive at the company.

**EDU_9**: It is nine o’clock.
**EDU_10**: Thank God, I am not late for work.

**EDU_11**: But the hamburger is cold,
**EDU_12**: order some take-away food is better, maybe.
Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

Stack, Buffer, Arcs set

Left-Arc(LA) \[\langle e|S, e'|B, Arcs\rangle \rightarrow \langle S, e'|B, Arcs \cup \{(e', e)\}\rangle\]

Right-Arc(RA) \[\langle e|S, e'|B, Arcs\rangle \rightarrow \langle e'|e|S, B, Arcs \cup \{(e, e')\}\rangle\]

Shift \[\langle S, e|B, Arcs\rangle \rightarrow \langle e|S, B, Arcs\rangle\]

Reduce \[\langle e|S, B, Arcs\rangle \rightarrow \langle S, B, Arcs\rangle\]
## Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

### Stack, Buffer, Arcs set

<table>
<thead>
<tr>
<th>Rule</th>
<th>Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left-Arc(LA)</strong></td>
<td>$$\langle e</td>
</tr>
<tr>
<td><strong>Right-Arc(RA)</strong></td>
<td>$$\langle e</td>
</tr>
<tr>
<td><strong>Shift</strong></td>
<td>$$\langle S, e</td>
</tr>
<tr>
<td><strong>Reduce</strong></td>
<td>$$\langle e</td>
</tr>
</tbody>
</table>
Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

**Stack, Buffer, Arcs set**

- **Left-Arc (LA)**: \( \langle e | S, e' | B, Arcs \rangle \rightarrow \langle S, e' | B, Arcs \cup \{(e', e)\} \rangle \)
- **Right-Arc (RA)**: \( \langle e | S, e' | B, Arcs \rangle \rightarrow \langle e | e | S, B, Arcs \cup \{(e, e')\} \rangle \)
- **Shift**: \( \langle S, e | B, Arcs \rangle \rightarrow \langle e | S, B, Arcs \rangle \)
- **Reduce**: \( \langle e | S, B, Arcs \rangle \rightarrow \langle S, B, Arcs \rangle \)
Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

**Stack, Buffer, Arcs set**

- **Left-Arc (LA)**: $\langle e | S, e' | B, Arcs \rangle \rightarrow \langle S, e' | B, Arcs \cup \{(e', e)\} \rangle$

- **Right-Arc (RA)**: $\langle e | S, e' | B, Arcs \rangle \rightarrow \langle e' | e | S, B, Arcs \cup \{(e, e')\} \rangle$

- **Shift**: $\langle S, e | B, Arcs \rangle \rightarrow \langle e | S, B, Arcs \rangle$

- **Reduce**: $\langle e | S, B, Arcs \rangle \rightarrow \langle S, B, Arcs \rangle$
Framework

Transition-based dependency parsing

Arc-eager algorithm (Nivre):

**Stack, Buffer, Arcs set**

Left-Arc(LA) \[ \langle e|S, e'|B, Arcs \rangle \rightarrow \langle S, e'|B, Arcs \cup \{(e', e)\} \rangle \]

Right-Arc(RA) \[ \langle e|S, e'|B, Arcs \rangle \rightarrow \langle e'|e|S, B, Arcs \cup \{(e, e')\} \rangle \]

Shift \[ \langle S, e|B, Arcs \rangle \rightarrow \langle e|S, B, Arcs \rangle \]

Reduce \[ \langle e|S, B, Arcs \rangle \rightarrow \langle S, B, Arcs \rangle \]
Arc-eager

**EDU_1**: President Bush insists
**EDU_2**: it would be a great tool
**EDU_3**: for curbing the budget deficit
**EDU_4**: and slicing the lard out of government programs.
**EDU_5**: He wants it now.

...  

**EDU_{32}**: Mr. Bush is considering simply declaring
**EDU_{33}**: that the Constitution gives him the power

...
Arc-eager

Transition Stack Buffer
[] [E₁, E₂, E₃, E₄, …]

EDU₁: President Bush insists
EDU₂: it would be a great tool
EDU₃: for curbing the budget deficit
EDU₄: and slicing the lard out of government programs.
EDU₅: He wants it now.
...
EDU₃₂: Mr. Bush is considering simply declaring
EDU₃₃: that the Constitution gives him the power
...

E₁ E₂ E₃ E₄ ...
**Arc-eager**

<table>
<thead>
<tr>
<th>Transition</th>
<th>Stack</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[E₁]</td>
<td>[E₁, E₂, E₃, E₄, …]</td>
</tr>
<tr>
<td>Shift</td>
<td></td>
<td>[E₂, E₃, E₄, …]</td>
</tr>
</tbody>
</table>

- **EDU₁**: President Bush insists
- **EDU₂**: it would be a great tool
- **EDU₃**: for curbing the budget deficit
- **EDU₄**: and slicing the lard out of government programs.
- **EDU₅**: He wants it now.
- ...
- **EDU₃₂**: Mr. Bush is considering simply declaring
- **EDU₃₃**: that the Constitution gives him the power
- ...
President Bush insists it would be a great tool for curbing the budget deficit and slicing the lard out of government programs. He wants it now.

Mr. Bush is considering simply declaring that the Constitution gives him the power...
Arc-eager

Transition  Stack  Buffer
Shift       [E₁]    [E₁, E₂, E₃, E₄, ...]
LA(Attribution)  []     [E₂, E₃, E₄, ...]
SH          [E₂]    [E₃, E₄, ...]

EDU₁: President Bush insists
EDU₂: it would be a great tool
EDU₃: for curbing the budget deficit
EDU₄: and slicing the lard out of government programs.
EDU₅: He wants it now.
...
EDU₃₂: Mr. Bush is considering simply declaring
EDU₃₃: that the Constitution gives him the power
...

Attribution

E₁  E₂  E₃  E₄  ...
Arc-eager

Transition       Stack       Buffer
Shift            [E₁]        [E₁, E₂, E₃, E₄, ...]
LA(Attribution)  []           [E₂, E₃, E₄, ...]
SH               [E₂]        [E₂, E₃, E₄, ...]
RA(Elaboration)  [E₂, E₃]   [E₃, E₄, ...]

EDU₁: President Bush insists
EDU₂: it would be a great tool
EDU₃: for curbing the budget deficit
EDU₄: and slicing the lard out of government programs.
EDU₅: He wants it now.

EDU₃₂: Mr. Bush is considering simply declaring
EDU₃₃: that the Constitution gives him the power

...
Arc-eager

Transition Stack Buffer
Shift [E₁] [E₁, E₂, E₃, E₄, ⋯]
LA (Attribution) [] [E₂, E₃, E₄, ⋯]
SH [E₂] [E₃, E₄, ⋯]
RA (Elaboration) [E₂, E₃] [E₄, ⋯]
RA (Joint) [E₂, E₃, E₄] [⋯]

EDU₁: President Bush insists
EDU₂: it would be a great tool
EDU₃: for curbing the budget deficit
EDU₄: and slicing the lard out of government programs.
EDU₅: He wants it now.
...
EDU₃₂: Mr. Bush is considering simply declaring
EDU₃₃: that the Constitution gives him the power
...

E₁ E₂ E₃ E₄ ⋯
President Bush insists it would be a great tool for curbing the budget deficit and slicing the lard out of government programs. He wants it now.

Mr. Bush is considering simply declaring that the Constitution gives him the power...
Arc-eager

<table>
<thead>
<tr>
<th>Transition</th>
<th>Stack</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shift</td>
<td>[E₁]</td>
<td>[E₁, E₂, E₃, E₄, ⋯]</td>
</tr>
<tr>
<td>LA(Attribution)</td>
<td>[]</td>
<td>[E₂, E₃, E₄, ⋯]</td>
</tr>
<tr>
<td>SH</td>
<td>[E₂]</td>
<td>[E₃, E₄, ⋯]</td>
</tr>
<tr>
<td>RA(Elaboration)</td>
<td>[E₂, E₃]</td>
<td>[E₄, ⋯]</td>
</tr>
<tr>
<td>RA(Joint)</td>
<td>[E₂, E₃, E₄]</td>
<td>⋯</td>
</tr>
</tbody>
</table>

EDU₁: President Bush insists
EDU₂: it would be a great tool
EDU₃: for curbing the budget deficit
EDU₄: and slicing the lard out of government programs.
EDU₅: He wants it now.
...
EDU₃₂: Mr. Bush is considering simply declaring
EDU₃₃: that the Constitution gives him the power
...
Model Overview

time $t$ transition state
Model Overview

time $t$ transition state

State Representation
Model Overview

**time $t$ transition state**

Transition(action-relation) distributions

State Representation

Memory network 1

Memory network 2
**Refined EDU**

```
BRefined

BRefined
VB
B_Coh

weighted sum

match

Memory network2

{ ...

wi

slot_j
```

**Word**

**POS**

**Position1**

- **Bi-LSTM**
- **Attention**

- **Bi-LSTM**
- **Position1**
Refined

EDU basic representation
Refined

EDU basic representation

Position in the sentence, paragraph and discourse
The diagram illustrates the process of refining EDUs (Educational Units) through a weighted sum of similarity scores. The weighted sum is calculated as:

\[ w_i = \frac{\exp(\lambda \cos(V_B, slot_i))}{\sum_j \exp(\lambda \cos(V_B, slot_j))} \]

where \( V_B \) is the vector representation of the EDU, and \( slot_i \) is the slot of the EDU in the discourse. The refined EDU (BRefined) is obtained by this weighted sum:

\[ B_{coh} = \sum_i w_i slot_i \]

The EDU basic representation is obtained through a memory network, which processes the word, POS, and position information using Bi-LSTM attention mechanisms. The position information includes the sentence, paragraph, and discourse levels.
**Refined EDU basic representation**

**Position in the sentence, paragraph and discourse**

\[
S_{\text{Coh}} = \sum_i w_i \cdot \text{slot}_i
\]

\[
w_i = \frac{\exp(\lambda \cos(V_S, \text{slot}_i))}{\sum_j \exp(\lambda \cos(V_S, \text{slot}_j))}
\]

**EDU basic representation**

**Position**
A and Position2

Top three transition information
Concatenate every transition’s embedding
A and Position2

Top three transition information
Concatenate every transition’s embedding

The spatial relationship between the top EDUs of S and B
• Same sentence
• Same paragraph
• Distance in paragraph
Transitions Sequence:
Shift, LA-attribute, SH, RA-elaboration, RA-joint, ⋯

**EDU₁**: President Bush insists

**EDU₂**: it would be a great tool

**EDU₃**: for curbing the budget deficit

**EDU₄**: and slicing the lard out of government programs.

**EDU₅**: He wants it now.

⋯

**EDU₃₂**: Mr. Bush is considering simply declaring

**EDU₃₃**: that the Constitution gives him the power

⋯
Experiment

Dataset:
RST Discourse Treebank
- 380 discourses
  - 312 training, 30 validation, 38 testing
- 111 relation types for fine-grained
- 19 relation types for coarse-grained
Experiment

Dataset:
RST Discourse Treebank
- 380 discourses
  - 312 training, 30 validation, 38 testing
- 111 relation types for fine-grained
- 19 relation types for coarse-grained

Evaluation metrics:
- UAS, LAS
Position features provide useful **structural clues** to our parser
Memory Network could **model the discourse cohesion info** such as lexical chains, topical infos so as to provide clues to our parser.

<table>
<thead>
<tr>
<th>Method</th>
<th>UAS</th>
<th>LAS (Fine)</th>
<th>LAS (Coarse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptron</td>
<td>0.5422</td>
<td>0.3231</td>
<td>0.3777</td>
</tr>
<tr>
<td>Basic (word+POS)</td>
<td>0.5588</td>
<td>0.3670</td>
<td>0.3985</td>
</tr>
<tr>
<td>Basic (word+POS+position)</td>
<td>0.5933</td>
<td>0.3832</td>
<td>0.4305</td>
</tr>
<tr>
<td>Main-full</td>
<td><strong>0.6197</strong></td>
<td><strong>0.3947</strong></td>
<td><strong>0.4445</strong></td>
</tr>
<tr>
<td>MST-full</td>
<td>0.7331</td>
<td>0.4309</td>
<td>0.4851</td>
</tr>
</tbody>
</table>
### Experiment (Cont.)

<table>
<thead>
<tr>
<th>Method</th>
<th>UAS</th>
<th>LAS (Fine)</th>
<th>LAS (Coarse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptron</td>
<td>0.5422</td>
<td>0.3231</td>
<td>0.3777</td>
</tr>
<tr>
<td>Basic (word+POS)</td>
<td>0.5588</td>
<td>0.367</td>
<td>0.3985</td>
</tr>
<tr>
<td>Basic (word+POS+position)</td>
<td>0.5933</td>
<td>0.3832</td>
<td>0.4305</td>
</tr>
<tr>
<td>Main-full</td>
<td>0.6197</td>
<td>0.3947</td>
<td>0.4445</td>
</tr>
<tr>
<td>MST-full</td>
<td><strong>0.7331</strong></td>
<td><strong>0.4309</strong></td>
<td><strong>0.4851</strong></td>
</tr>
</tbody>
</table>

MST-full (graph-based) can directly analyze the relationship between any EDU pairs.
Conclusions & Future work

Conclusions:

We propose to utilize memory networks to model discourse cohesion automatically.

• Capture the topic change or lexical chains within a discourse
Conclusions & Future work

Conclusions:

We propose to utilize memory networks to model discourse cohesion automatically.

- Capture the topic change or lexical chains within a discourse

Improve the discourse parsing performance
Conclusions & Future work

Conclusions:

We propose to utilize memory networks to model discourse cohesion automatically.

- Capture the topic change or lexical chains within a discourse

Improve the discourse parsing performance

Future work:

Apply our method on the graph-based parsing system

Optimize memory network structure
Thanks