Memory-Enhanced Models for Discourse Understanding

COMP90042 Web Search and Text Analysis
Guest Lecture

Fei Liu

School of Computing and Information Systems
The University of Melbourne

May 28th, 2019
Table of Contents

1. What is Discourse
2. Discourse-related Tasks
3. Models for Discourse Understanding
4. Conclusion
What is Discourse

Discourse-related Tasks

Models for Discourse Understanding

Conclusion
Discourse

Discourse: a coherent, structured group of sentences (utterances)

Example

Yesterday, Ted was late for work. [It all started when his car wouldn’t start. He first tried to jump start it with a neighbour’s help, but that didn’t work.] [So he decided to take public transit. He walked 15 minutes to the tram stop. Then he waited for another 20 minutes, but the tram didn’t come. The tram drivers were on strike that morning.] [So he walked home and got his bike out of the garage. He started riding but quickly discovered he had a flat tire. He walked his bike back home. He looked around but his wife had cleaned the garage and he couldn’t find the bike pump.] He started walking, and didn’t arrive until lunchtime.\(^a\)

\(^a\)Example from WSTA L20
Discourse-related Tasks
Sentence-level Discourse Understanding Tasks

Coreference resolution: grouping all expressions referring to the same entity into the same cluster (implicitly requires the detection of entities, either do entity recognition in a pipeline or jointly with coreference resolution)

Example

He first tried to jump start it with a neighbour’s help, but that didn’t work.
Sentence-level Discourse Understanding Tasks

Coreference resolution: grouping all expressions referring to the same entity into the same cluster (implicitly requires the detection of entities, either do entity recognition in a pipeline or jointly with coreference resolution)

Example

It all started when his car wouldn’t start. He first tried to jump start it with a neighbour’s help, but that didn’t work.
Sentence-level Discourse Understanding Tasks

**Winograd**: pronoun disambiguation, requiring a deep semantic understanding of text (Levesque et al., 2012)

Example

The woman held the girl against her **chest**.
The woman held the **girl** against her **will**.
Sentence-level Discourse Understanding Tasks

**Winograd**: pronoun disambiguation, requiring a deep semantic understanding of text (Levesque et al., 2012)

**Example**

The *city councilmen* refused the demonstrators a permit because they *feared* violence.
The city councilmen refused the *demonstrators* a permit because they *advocated* violence.

Challenging task with a success rate of $\approx 70\%$ by recent works (Radford et al., 2019, Kocijan et al., 2019)
Plenty of room for improvement
Sentence-level Discourse Understanding Tasks

Recent development on Winograd

- Random
- Liu et al. (2016)
- Trinh et al. (2018)
- Nangia et al. (2018)
- Radford et al. (2019)
- Kocijan et al. (2019)
Segment-level Discourse Understanding Tasks

**Discourse segmentation**: identifying the boundaries between different segments of text

Example

[It all started when *his car* wouldn’t *start*. He first tried to *jump start it* with a neighbour’s help, but *that didn’t work.*] [So he decided to take *public transit*. He walked 15 minutes to the *tram stop*. Then *he* waited for another 20 minutes, but the *tram* didn’t come. The *tram drivers* were on strike *that morning*.]
Short-story Understanding Tasks

**Story Cloze Test**: predicting the most coherent ending options to a given 4-sentence short story (Mostafazadeh et al., 2016)

**Example**

**Story**: Sam loved his old belt. He matched it with everything. Unfortunately he gained too much weight. It became too small.

**Coherent ending**: Sam went on a diet. ✓

**Incoherent ending**: Sam was happy. ✗

**Example**

**Story**: Rick fell while hiking in the woods. He was terrified! He thought he had fallen into a patch of poison ivy. Then he used his nature guide to identify the plant.

**Coherent ending**: He was relieved to find out he was wrong. ✓

**Incoherent ending**: Rick was soaking wet from falling in the pond. ✗
Story Understanding: a toy dataset bAbI

bAbI: reasoning-focused question answering (Weston et al., 2016)

Example

<table>
<thead>
<tr>
<th>#</th>
<th>Story</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jeff went to the kitchen.</td>
</tr>
<tr>
<td>2</td>
<td>Mary travelled to the hallway.</td>
</tr>
<tr>
<td>3</td>
<td>Jeff picked up the milk.</td>
</tr>
<tr>
<td>4</td>
<td>Jeff travelled to the bedroom.</td>
</tr>
<tr>
<td>5</td>
<td>Jeff left the milk there.</td>
</tr>
<tr>
<td>6</td>
<td>Jeff went to the bathroom.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where is the milk now?</td>
<td>bedroom</td>
</tr>
<tr>
<td>Where is Jeff?</td>
<td>bathroom</td>
</tr>
<tr>
<td>Where was Jeff before the bedroom?</td>
<td>kitchen</td>
</tr>
</tbody>
</table>

Table: Key pieces of evidence for the first question are underlined, with distractors marked with dashed underline.
**SQuAD**: reading comprehension with the answer being a continuous span of text in the given document (Rajpurkar et al., 2016, 2018)

**Example**

**Document**: Victoria (abbreviated as Vic) is a state in the **south-east** of Australia. Victoria is Australia’s most densely populated state and its second-most populous state overall. Most of its population is concentrated in the area surrounding Port Phillip Bay, which includes the metropolitan area of its capital and largest city, Melbourne, which is Australia’s second-largest city. Geographically the smallest state on the Australian mainland, Victoria is bordered by Bass Strait and Tasmania to the south, New South Wales to the north, the Tasman Sea to the east, and South Australia to the west.

**Question**: Where in Australia is Victoria located?

**Answer**: south-east
Document-level Reading Comprehension: SQuAD

**SQuAD**: reading comprehension with the answer being a continuous span of text in the given document (Rajpurkar et al., 2016, 2018)

**Example**

**Document**: Victoria (abbreviated as Vic) is a state in the south-east of Australia. Victoria is Australia’s **most densely populated** state and its second-most populous state overall. Most of its population is concentrated in the area surrounding Port Phillip Bay, which includes the metropolitan area of its capital and largest city, Melbourne, which is Australia’s second-largest city. Geographically the smallest state on the Australian mainland, Victoria is bordered by Bass Strait and Tasmania to the south, New South Wales to the north, the Tasman Sea to the east, and South Australia to the west.

**Question**: How does Victoria rank as to population density?

**Answer**: most densely populated
**Document-level Reading Comprehension: SQuAD**

**SQuAD**: reading comprehension with the answer being a continuous span of text in the given document (Rajpurkar et al., 2016, 2018)

**Example**

**Document**: Victoria (abbreviated as Vic) is a state in the south-east of Australia. Victoria is Australia’s most densely populated state and its second-most populous state overall. Most of its population is concentrated in the area surrounding Port Phillip Bay, which includes the metropolitan area of its capital and largest city, Melbourne, which is Australia’s second-largest city. Geographically the smallest state on the Australian mainland, Victoria is bordered by Bass Strait and Tasmania to the south, New South Wales to the north, the Tasman Sea to the east, and South Australia to the west.

**Question**: How does Melbourne rank as to population?

**Answer**: <No Answer>
Multi-document Reading Comprehension: QAngaroo

**QAngaroo**: multi-document comprehension (Welbl et al., 2017)

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Big Oak Tree State Park</strong> is a state-owned nature preserve ... in the Mississippi Alluvial Plain portion of the <strong>Gulf Coastal Plain</strong>.</td>
</tr>
<tr>
<td><strong>The Gulf Coastal Plain</strong> extends around the Gulf of Mexico in the <strong>Southern United States</strong> ...</td>
</tr>
<tr>
<td><strong>The Southern United States</strong>, commonly referred to as the American South, Dixie, or simply the South, is a region of the <strong>United States of America</strong>.</td>
</tr>
</tbody>
</table>

**Question**: Where is **Big Oak Tree State Park** located?

**Answer**: United States of America
Dialog State Tracking

**Dialog state tracking**: maintaining up-to-date slot values regarding dialog states

### Example

<table>
<thead>
<tr>
<th>User</th>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hello and welcome</td>
<td></td>
</tr>
<tr>
<td>What kind of food would you like?</td>
<td></td>
</tr>
<tr>
<td>Moderately priced Swedish food</td>
<td></td>
</tr>
</tbody>
</table>

**food**: Swedish, price range: moderate, area: none

**Table**: An example from DSTC-2 (Henderson et al., 2014)
Dialog State Tracking

**Dialog state tracking**: maintaining up-to-date slot values regarding dialog states

**Example**

<table>
<thead>
<tr>
<th>User</th>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hello and welcome</td>
<td>Moderately priced Swedish food</td>
</tr>
<tr>
<td>What kind of food would you like?</td>
<td>Sorry there is no Swedish restaurant in the moderate price range</td>
</tr>
<tr>
<td>How about Asian food?</td>
<td>food: Asian, price range: moderate, area: none</td>
</tr>
</tbody>
</table>

**Table**: An example from DSTC-2 (Henderson et al., 2014)
Models for Discourse Understanding
**Memory Networks**

**Memory Networks**: progressively incorporating evidence from the previous reasoning hop (Sukhbaatar et al., 2015)

![Illustration of a memory network with a single memory hop.](image)

**Figure**: Illustration of a memory network with a single memory hop.

\[ p_i = \text{softmax}(u \cdot m_i), \quad o = \sum_{i=1}^{m} p_i m_i, \quad \hat{a} = \text{softmax}(W(o + u)) \]
Memory Networks

**Memory Networks**: progressively incorporating evidence from the previous reasoning hop (Sukhbaatar et al., 2015)

![Diagram of a memory network with multiple memory hops](image)

**Figure**: Illustration of a memory network with multiple memory hops.

\[
p_i^{(k)} = \text{softmax}(u^{(k)} \cdot m_i^{(k)}), \quad o^{(k)} = \sum_{i=1}^{m} p_i^{(k)} m_i^{(k)}
\]

\[
u^{(k+1)} = u^{(k)} + o^{(k)}, \quad \hat{a} = \text{softmax}(W(o + u))
\]
Memory Networks

<table>
<thead>
<tr>
<th>Story</th>
<th>Support</th>
<th>Memory Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeff went to the kitchen.</td>
<td></td>
<td>Hop 1  Hop 2  Hop 3</td>
</tr>
<tr>
<td>Mary travelled to the hallway.</td>
<td></td>
<td>0.00  0.00  0.00</td>
</tr>
<tr>
<td>Jeff picked up the milk.</td>
<td>yes</td>
<td>0.82  0.00  0.00</td>
</tr>
<tr>
<td>Jeff travelled to the bedroom.</td>
<td>yes</td>
<td>0.00  1.00  0.00</td>
</tr>
<tr>
<td>Jeff left the milk there.</td>
<td>yes</td>
<td>0.18  0.00  1.00</td>
</tr>
<tr>
<td>Jeff went to the bathroom.</td>
<td></td>
<td>0.00  0.00  0.00</td>
</tr>
</tbody>
</table>

Question: Where is the milk now?
Answer: bedroom, prediction: bedroom

Table: An example of the attention weights of a 3-hop memory network trained on task 5 (3 argument relations) of the bAbI dataset. True supporting sentences are marked “yes” in the support column.
Dynamic Memory Chains

**Recurrent Entity Networks:** keeping track of entity states with external memory chains (Henaff et al., 2017)
Dynamic Memory Chains

1. Jeff went to the kitchen. ((1) Jeff in kitchen)
Dynamic Memory Chains

1. Jeff went to the kitchen. ((1) Jeff in kitchen)
2. Mary travelled to the hallway. ((1) Jeff in kitchen, (2) Mary in hallway)
Dynamic Memory Chains

1. Jeff went to the kitchen. (1) Jeff in kitchen
2. Mary travelled to the hallway. (1) Jeff in kitchen, (2) Mary in hallway
3. Jeff picked up the milk. (1) Jeff in kitchen carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in kitchen
Dynamic Memory Chains

1. Jeff went to the kitchen. ((1) Jeff in kitchen)
2. Mary travelled to the hallway. ((1) Jeff in kitchen, (2) Mary in hallway)
3. Jeff picked up the milk. ((1) Jeff in kitchen carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in kitchen)
4. Jeff travelled to the bedroom. ((1) Jeff in bedroom carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in bedroom)
Dynamic Memory Chains

1. Jeff went to the kitchen. ((1) Jeff in kitchen)
2. Mary travelled to the hallway. ((1) Jeff in kitchen, (2) Mary in hallway)
3. Jeff picked up the milk. ((1) Jeff in kitchen carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in kitchen)
4. Jeff travelled to the bedroom. ((1) Jeff in bedroom carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in bedroom)
5. Jeff left the milk there. ((1) Jeff in bedroom carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in bedroom)
Dynamic Memory Chains

1. Jeff went to the kitchen. ((1) Jeff in kitchen)
2. Mary travelled to the hallway. ((1) Jeff in kitchen, (2) Mary in hallway)
3. Jeff picked up the milk. ((1) Jeff in kitchen carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in kitchen)
4. Jeff travelled to the bedroom. ((1) Jeff in bedroom carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in bedroom)
5. Jeff left the milk there. ((1) Jeff in bedroom carrying milk, (2) Mary in hallway, (3) milk carried by Jeff in bedroom)
6. Jeff went to the bathroom. ((1) Jeff in bathroom, (2) Mary in hallway, (3) milk in bedroom)
### Dynamic Memory Chains

<table>
<thead>
<tr>
<th>#</th>
<th>Story</th>
<th>Jeff</th>
<th>Mary</th>
<th>Milk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jeff went to the kitchen.</td>
<td>in kitchen</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>Mary travelled to the hallway.</td>
<td>in kitchen</td>
<td>in hallway</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>Jeff picked up the milk.</td>
<td>in kitchen</td>
<td>in hallway</td>
<td>in kitchen carried by Jeff</td>
</tr>
<tr>
<td>4</td>
<td>Jeff travelled to the bedroom.</td>
<td>in bedroom carrying milk</td>
<td>in hallway</td>
<td>in bedroom carried by Jeff</td>
</tr>
<tr>
<td>5</td>
<td>Jeff left the milk.</td>
<td>in bedroom carrying milk</td>
<td>in hallway</td>
<td>in bedroom carried by Jeff</td>
</tr>
<tr>
<td>6</td>
<td>Jeff went to the bathroom.</td>
<td>in bedroom</td>
<td>in hallway</td>
<td>in bedroom</td>
</tr>
</tbody>
</table>

**Table:** Dynamic memory chains keeping track of entities: Jeff, Mary and milk. The states of such entities are updated as new input is processed.
Dynamic Memory Chains for Narrative Understanding

**Story Cloze Test**: predicting the most coherent ending options to a given 4-sentence short story (Mostafazadeh et al., 2016)

**Example**

**Story**: Rick fell while hiking in the woods. He was terrified! He thought he had fallen into a patch of poison ivy. Then he used his nature guide to identify the plant.

**Coherent ending**: He was relieved to find out he was wrong. ✓

**Incoherent ending**: Rick was soaking wet from falling in the pond. X
Dynamic Memory Chains for Narrative Understanding

**Key motivation:** understanding a story from three perspectives: (1) event sequence, (2) sentiment trajectory, (3) topic consistency.
Dynamic Memory Chains for Narrative Understanding

Without Semantic Supervision

With Semantic Supervision

Event Key  Sentiment Key  Topic Key  Free Key

- Event Word  - Sentiment Word  - Topic Word
Coreference Resolution

**RefReader**: online text processing with a fixed-size working memory (Liu et al., 2019)

**Figure**: A referential reader with two memory cells. Overwrite and update are indicated by $o_t^{(i)}$ and $u_t^{(i)}$; in practice, these operations are continuous gates. Thickness and color intensity of edges between memory cells at neighboring steps indicate memory salience; $\times$ indicates an overwrite.
Coreference Resolution

**RefReader**: compute token pair-wise coreferential probability: token at time $t_2$ referring to that at $t_1$ is defined as

$$\hat{\psi}_{t_1, t_2} = \sum_{i=1}^{N} (u_{t_1}^{(i)} + o_{t_1}^{(i)}) \times u_{t_2}^{(i)} \times \prod_{t=t_1+1}^{t_2} (1 - o_t^{(i)})$$

- update or overwrite at time $t_1$
- update at time $t_2$
- not overwritten in $[t_1 + 1, t_2]$
Coreference Resolution

**RefReader**: target coreference matrix:

<table>
<thead>
<tr>
<th></th>
<th>Barack</th>
<th>Obama</th>
<th>told</th>
<th>Xi</th>
<th>Jinping</th>
<th>his</th>
<th>concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barack</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Obama</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>told</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Xi</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jinping</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>his</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>concerns</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table**: Example of the target coreference matrix with light and dark gray highlighting self-link and pronoun coreferential cells.
Conclusion
Conclusion

1 Discourse-related Tasks
   1.1 Sentence-level Tasks
   1.2 Short Story-level Tasks
   1.3 Document-level Tasks
   1.4 Dialogue Tasks

2 Memory-enhanced Models for Discourse Understanding
   2.1 Memory Networks
   2.2 Dynamic Memory Chains
   2.3 REFReader for Coreference Resolution
## References