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

What is Discourse

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

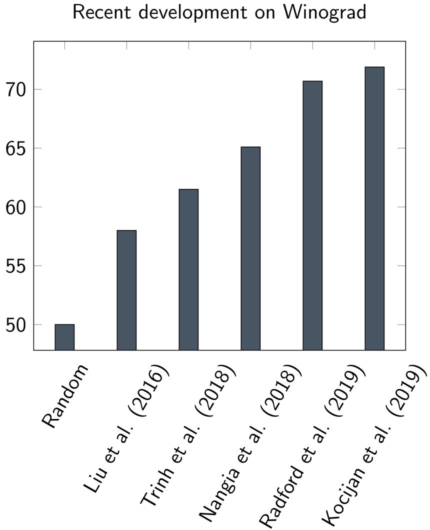
Winograd: pronoun disambiguation, requiring a deep semantic understanding of text (Levesque et al., 2012)

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 ≈ 70% by recent works (Radford et al., 2019, Kocijan et al., 2019)
Plenty of room for improvement

Sentence-level Discourse Understanding Tasks



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

#	Story	
1	Jeff went to the kitchen.	
2	Mary travelled to the hallway.	
3	Jeff picked up the milk.	
4	Jeff travelled to the bedroom.	
5	Jeff left the milk there.	
6	Jeff went to the bathroom.	
Question		**Answer**
Where is the milk now?		bedroom
Where is Jeff?		bathrom
Where was Jeff before the bedroom?		kitchen

Table: Key pieces of evidence for the first question are underlined, with distractors marked with dashed underline.

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: 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)

Example

Big Oak Tree State Park is a state-owned nature preserve . . . in the Mississippi Alluvial Plain portion of the Gulf Coastal Plain.

The Gulf Coastal Plain extends around the Gulf of Mexico in the Southern United States . . .

The Southern United States, commonly referred to as the American South, Dixie, or simply the South, is a region of the United States of America.

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

User	Agent
	Hello and welcome
	What kind of food would you like?
Moderately priced Swedish food	
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

User	Agent
	Hello and welcome
	What kind of food would you like?
Moderately priced Swedish food	
Sorry there is no Swedish restaurant in the moderate price range	
How about Asian food?	
food: Asian, price range: moderate, area: none	

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)

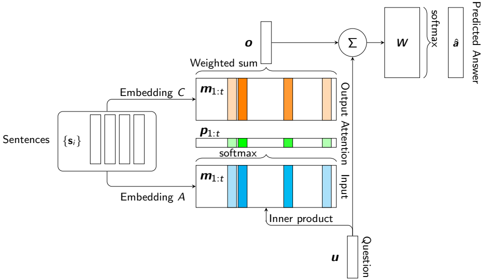


Figure: Illustration of a memory network with a single memory hop.

$$p_i = \text{softmax}(\boldsymbol{u} \cdot \boldsymbol{m}_i), \qquad \boldsymbol{o} = \sum_{i=1}^{m} p_i \boldsymbol{m}_i, \qquad \hat{\boldsymbol{a}} = \text{softmax}(\boldsymbol{W}(\boldsymbol{o} + \boldsymbol{u}))$$

Memory Networks

MEMORY NETWORKS: progressively incorporating evidence from the previous reasoning hop (Sukhbaatar et al., 2015)

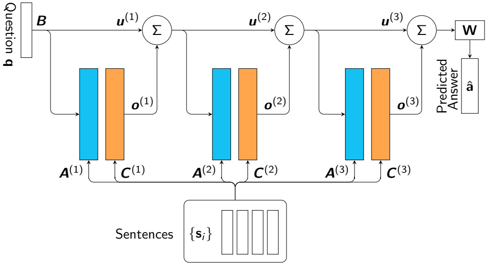


Figure: Illustration of a memory network with multiple memory hops.

$$p_i^{(k)} = \text{softmax}(\boldsymbol{u}^{(k)} \cdot \boldsymbol{m}_i^{(k)}), \qquad \boldsymbol{o}^{(k)} = \sum_{i=1}^{m} p_i^{(k)} \boldsymbol{m}_i^{(k)}$$

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \boldsymbol{o}^{(k)}, \qquad \hat{\boldsymbol{a}} = \text{softmax}(\boldsymbol{W}(\boldsymbol{o} + \boldsymbol{u}))$$

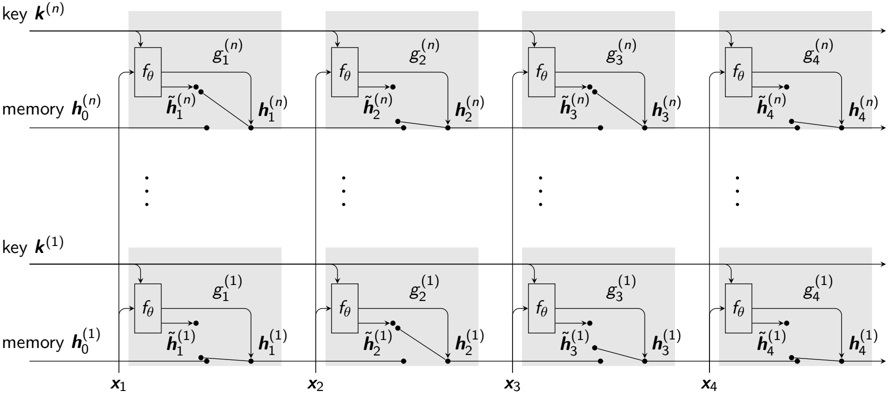
Memory Networks

Story	Support	Memory Network Hop 1	Hop 2	Hop 3
Jeff went to the kitchen.		0.00	0.00	0.00
Mary travelled to the hallway.		0.00	0.00	0.00
Jeff picked up the milk.	yes	0.82	0.00	0.00
Jeff travelled to the bedroom.	yes	0.00	1.00	0.00
Jeff left the milk there.	yes	0.18	0.00	1.00
Jeff went to the bathroom.		0.00	0.00	0.00
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

#	Story	Memory Chains Jeff	Memory Chains Mary	Memory Chains Milk
1	Jeff went to the kitchen.	in kitchen	–	–
2	Mary travelled to the hallway.	in kitchen	in hallway	–
3	Jeff picked up the milk.	in kitchen carrying milk	in hallway	in kitchen carried by Jeff
4	Jeff travelled to the bedroom.	in bedroom carrying milk	in hallway	in bedroom carried by Jeff
5	Jeff left the milk.	in bedroom ~~carrying milk~~	in hallway	in bedroom ~~carried by Jeff~~
6	Jeff went to the bathroom.	in bedroom	in hallway	in bedroom

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

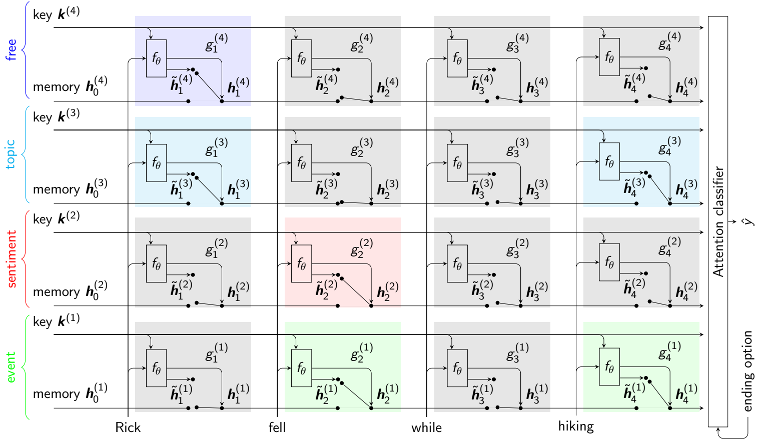
Story Cloze Test: predicting the most coherent ending options to a given 4-sentence short story (Mostafazadeh et al., 2016)

Example

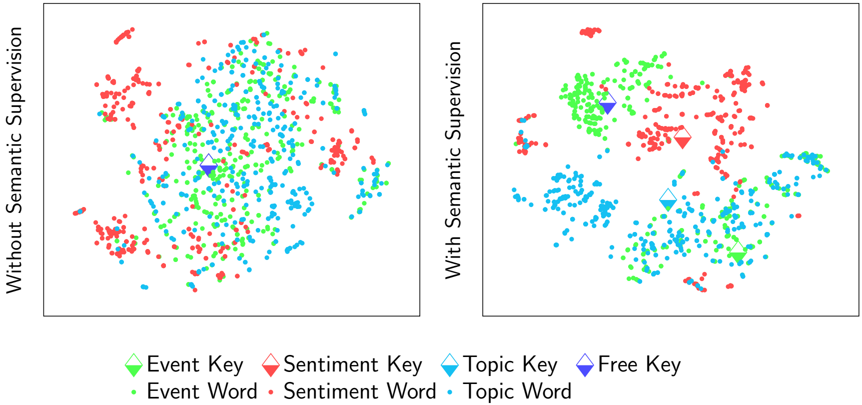
Story: Rick fell while hiking in the woods. He was terrified! He thought he had fallen into a patch of poison ivy. Then he used his nature guide to identify the plant.
Coherent ending: He was relieved to find out he was wrong. ✔
Incoherent ending: Rick was soaking wet from falling in the pond. ✘

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



Coreference Resolution

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

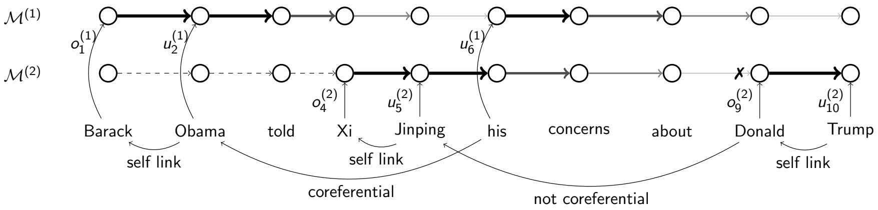


Figure: A referential reader with two memory cells. Overwrite and update are indicated by $o_t^{(i)}$ and $u_t^{(i)}$; in practice, these operations are continuous gates. Thickness and color intensity of edges between memory cells at neighboring steps indicate memory salience; ✗ indicates an overwrite.

Coreference Resolution

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

$$
\begin{aligned}
\hat{\psi}_{t_1,t_2} = \sum_{i=1}^{N} (u_{t_1}^{(i)} + o_{t_1}^{(i)}) & \quad \text{update or overwrite at time } t_1 \\
\times u_{t_2}^{(i)} & \quad \text{update at time } t_2 \\
\times \prod_{t=t_1+1}^{t_2} (1 - o_t^{(i)}) & \quad \text{not overwritten in } [t_1+1, t_2]
\end{aligned}
$$

Coreference Resolution

RefReader: target coreference matrix:

	Barack	Obama	told	Xi	Jinping	his	concerns
Barack	0	1	0	0	0	1	0
Obama	–	0	0	0	0	1	0
told	–	–	0	0	0	0	0
Xi	–	–	–	0	1	0	0
Jinping	–	–	–	–	0	0	0
his	–	–	–	–	–	0	0
concerns	–	–	–	–	–	–	0

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

Mikael Henaff, Jason Weston, Arthur Szlam, Antoine Bordes, and Yann LeCun. 2017. Tracking the world state with recurrent entity networks. In *Proceedings of the 5th International Conference on Learning Representations*, Toulon, France.

Matthew Henderson, Blaise Thomson, Jason D. Williams. 2014. The Second Dialog State Tracking Challenge. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 263–272, Philadelphia, USA.

Hector Levesque, Ernest Davis, and Leora Morgenstern. 2012. The winograd schema challenge. In *Proceedings of the Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*.

Fei Liu, Luke Zettlemoyer and Jacob Eisenstein (to appear) The Referential Reader: A RecurrentEntity Network for Anaphora Resolution. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy.

Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. 2016. A corpus and cloze evaluation for deeper understanding of commonsense stories. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 839–849, San Diego, USA.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang. 2016. SQuAD: 100,000+ Questions for Machine Comprehension of Text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, USA.

Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. Know What You Don't Know: Unanswerable Questions for SQuAD. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 784–789, Melbourne, Australia

Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. 2015. End-to-end memory networks. In *Proceedings of Advances in Neural Information Processing Systems*. Montréal, Canada, pages 2440–2448.

Johannes Welbl, Pontus Stenetorp, and Sebastian Riedel. 2018. Constructing Datasets for Multi-hop Reading Comprehension Across Documents. In *Transactions of the Association for Computational Linguistics*.

Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. 2016. Towards AI-complete question answering: A set of prerequisite toy tasks. In *Proceedings of the 4th International Conference on Learning Representations*, San Juan, Puerto Rico.