End-to-End Memory Networks

Paper by: S. Sukhbaatar et al.
Presentation by: Marina Samuel

Overview

- Problem: Text + Question => Answer
- Solution:
    - Recurrent attention model
    - Memory network trained end-to-end
    - Weakly supervised
- Solution extends to sentence generation
    - Performs better than previous comparable RNNs and LSTMs

Main Motivation

- Answer questions about a body of text
  - Try to sound like a human
  - Turing test
- Perform better than previous approaches

Statement of the Problem

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary?

Task 3: Three Supporting Facts
John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen?

Task 17: Positional Reasoning
The triangle is to the right of the blue square.
The red square is on top of the blue square.
The red sphere is to the right of the blue square.
Is the red sphere to the right of the blue square?
Is the red square to the left of the triangle?

Task 19: Path Finding
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?

- Idea: test in “units” - small, one word answers,
- Word order and sentence order matter

[J. Weston, A. Bordes, S. Chopra, et. al., “Towards AI Complete Question Answering: A Set of Prerequisite Toy Tasks”, ICLR 2016]

Statement of the Problem

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 17: Positional Reasoning

The triangle is to the right of the blue square.
The red square is on top of the blue square.
The red sphere is to the right of the blue square.
Is the red sphere to the right of the blue square?
Is the red square to the left of the triangle?

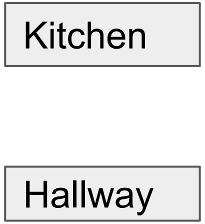
Task 3: Three Supporting Facts

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 19: Path Finding

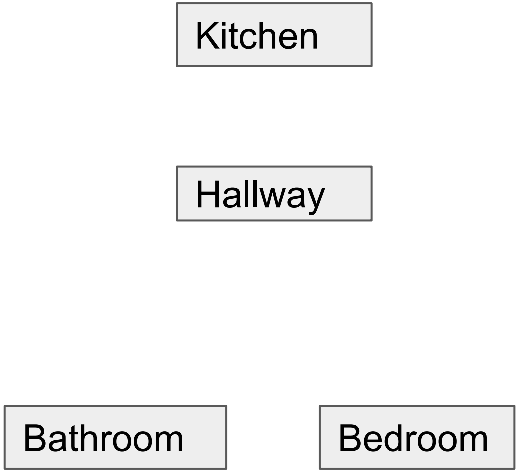
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?

- Idea: test in “units” - small, one word answers,
- Word order and sentence order matter



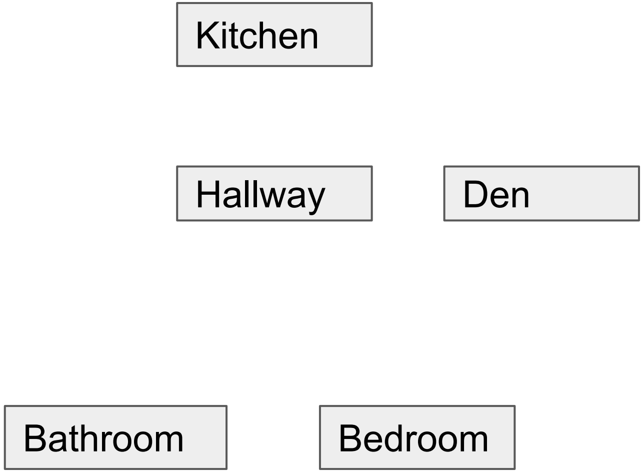
Task 19: Path Finding

The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?



Task 19: Path Finding

The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?

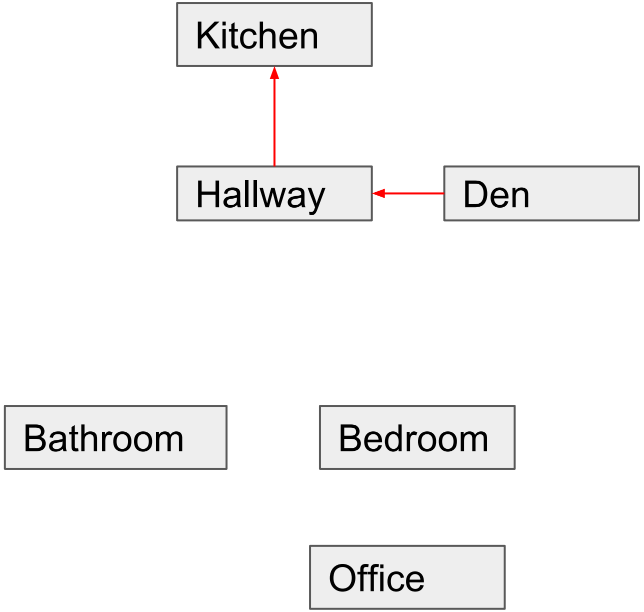


Task 19: Path Finding

The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?

Task 19: Path Finding

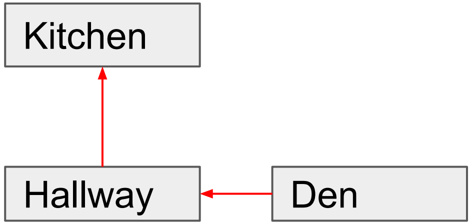
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?



Task 19: Path Finding

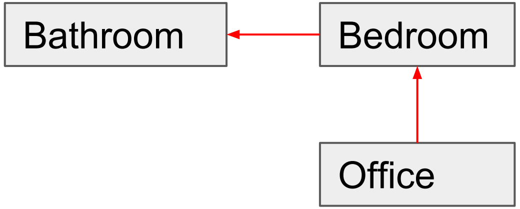
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?

Den -> Kitchen
West, North

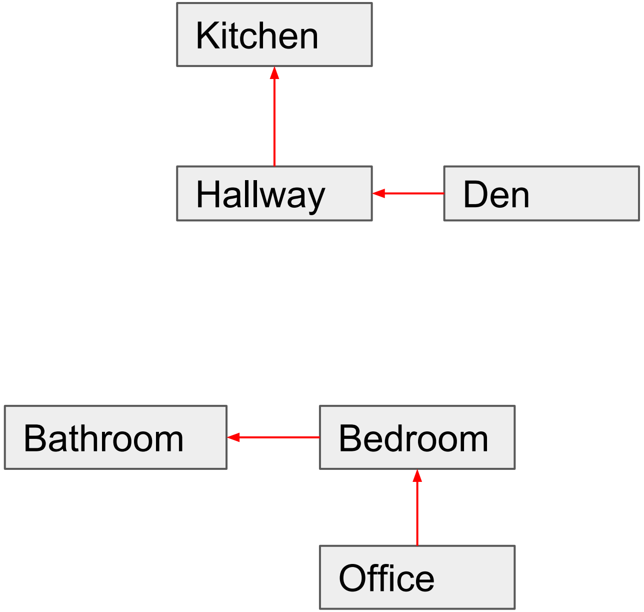


Task 19: Path Finding

The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?



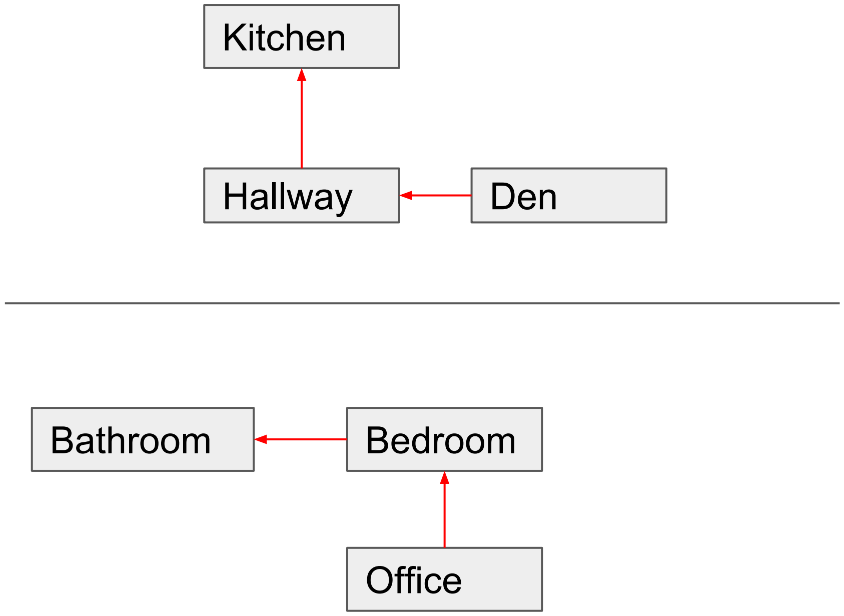
Office -> Bathroom
North, West



Task 19: Path Finding

The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen?
How do you go from office to bathroom?

How do you go from the kitchen to the bedroom?



How do you go from the kitchen to the bedroom?

Statement of the Problem

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 3: Three Supporting Facts

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 17: Positional Reasoning

The triangle is to the right of the blue square.
The red square is on top of the blue square.
The red sphere is to the right of the blue square.
Is the red sphere to the right of the blue square? A:yes
Is the red square to the left of the triangle? A:yes

Task 19: Path Finding

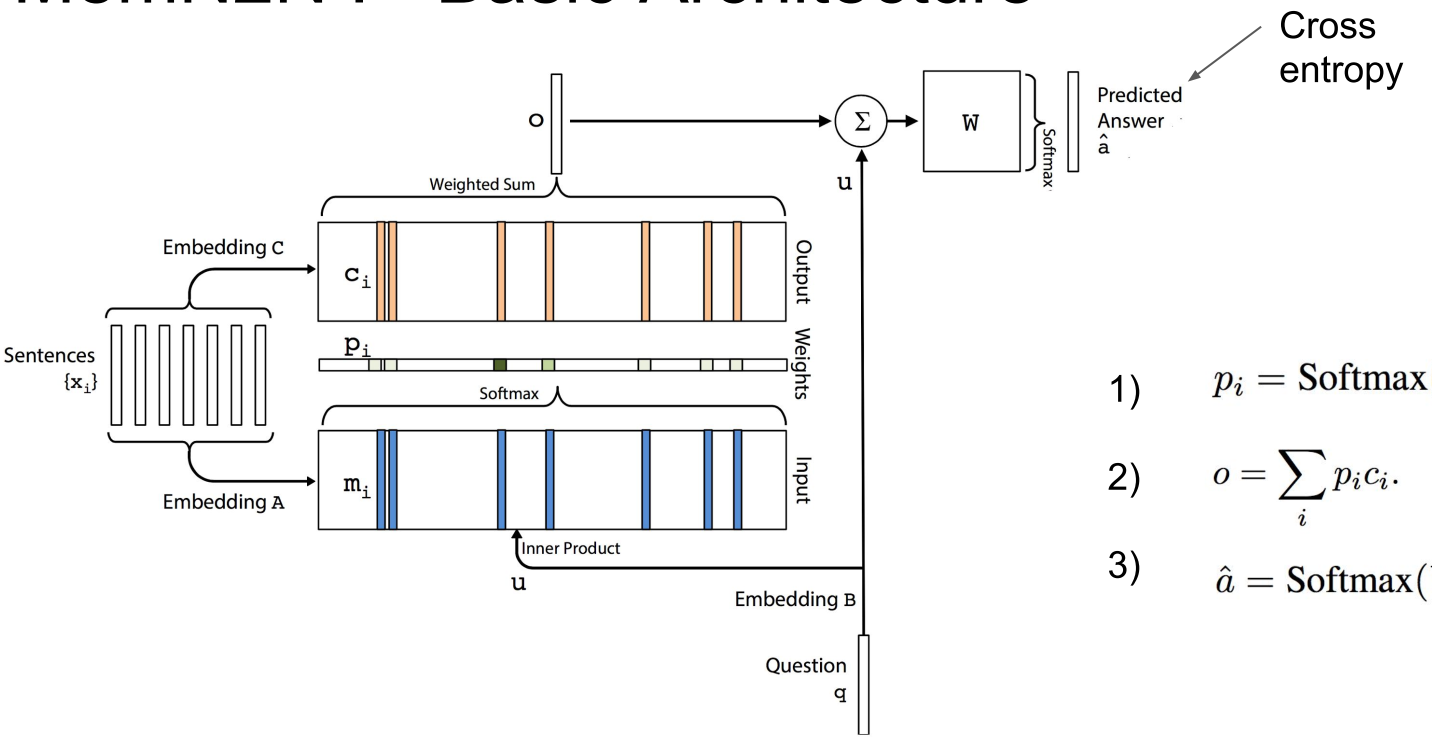
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen? A: west, north
How do you go from office to bathroom? A: north, west

- What assumptions are being made?
    - ordering
    - there is an answer (1 of 20)
    - nothing else happens

Memory Networks I

- Goal is to make use of memory in a neural net
- Compare to LSTM:
  - hidden state in each cell
  - Smaller amount of memory used
  - Changes easily over time

MemN2N I - Basic Architecture



1) $p_i = \mathrm{Softmax}(u^T m_i)$. ATTENTION!

2) $o = \sum_i p_i c_i.$

3) $\hat{a} = \mathrm{Softmax}(W(o+u))$

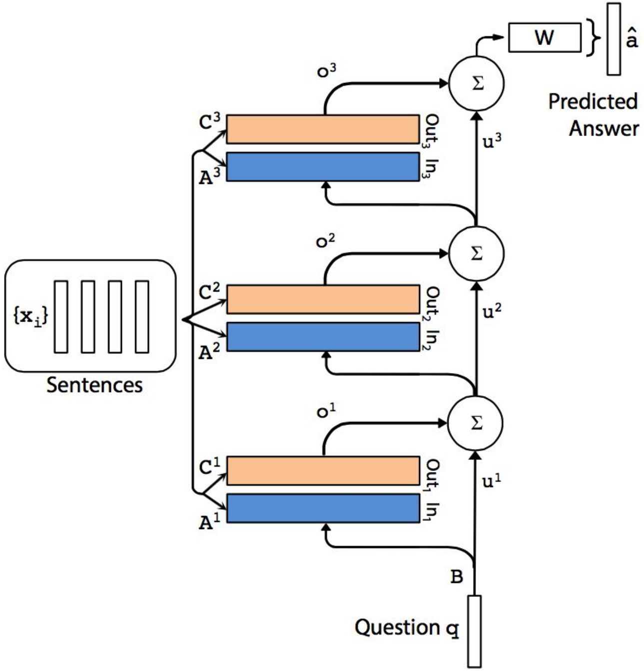
- Embedding matrices A, B, C, W are learned during training

[S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus, “End-to-End Memory Networks”, Nov 2015]

MemN2N II - Approaches to Embedding

- Bag of Words (BoW)
  - $m_i = \sum_j Ax_{ij}$
- Position Encoding (PE)
  - $m_i = \sum_j l_j \cdot Ax_{ij}$
- Temporal Encoding (TE)
  - $m_i = \sum_j Ax_{ij} + T_A(i)$
- Random Noise (RN)
  - For regularizing T_A

MemN2N III - Multi-hop Architectures (with k hops)



Types of Weight-Tying:

Adjacent:

- $A^{k+1} = C^k$
- $W^T = C^K$
- $B = A$

Layer-wise (RNN-Like):

- $A^1 = A^2 = ... = A^K$
- $C^1 = C^2 = ... = C^K$

[S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus, “End-to-End Memory Networks”, Nov 2015]

Attention During Hops

Story (1: 1 supporting fact)	Support	Hop 1	Hop 2	Hop 3
Daniel went to the bathroom.		0.00	0.00	0.03
Mary travelled to the hallway.		0.00	0.00	0.00
John went to the bedroom.		0.37	0.02	0.00
John travelled to the bathroom.	yes	0.60	0.98	0.96
Mary went to the office.		0.01	0.00	0.00
Where is John? Answer: bathroom Prediction: bathroom				

Story (2: 2 supporting facts)	Support	Hop 1	Hop 2	Hop 3
John dropped the milk.		0.06	0.00	0.00
John took the milk there.	yes	0.88	1.00	0.00
Sandra went back to the bathroom.		0.00	0.00	0.00
John moved to the hallway.	yes	0.00	0.00	1.00
Mary went back to the bedroom.		0.00	0.00	0.00
Where is the milk? Answer: hallway Prediction: hallway				

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow Prediction: yellow				

Story (18: size reasoning)	Support	Hop 1	Hop 2	Hop 3
The suitcase is bigger than the chest.	yes	0.00	0.88	0.00
The box is bigger than the chocolate.		0.04	0.05	0.10
The chest is bigger than the chocolate.	yes	0.17	0.07	0.90
The chest fits inside the container.		0.00	0.00	0.00
The chest fits inside the box.		0.00	0.00	0.00
Does the suitcase fit in the chocolate? Answer: no Prediction: no				

[S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus, “End-to-End Memory Networks”, Nov 2015]

MemN2N IV - Multi-hop vs. RNN

- Can think of u vector as hidden state
- Recurrence is: $u^{k+1} = u^k + o^k$
- Unlike RNN, there is no concrete intermediate output
    - e.g. not predicting a word at an intermediate state
    - u^{k+1} has no human-understandable meaning

MemN2N V - Sentence Generation

- Same multi-hop architecture
- Now embed words individually instead of sentences
- There is no question so u is fixed as constant vector, 0.1
- Layer-wise weight sharing
- Performs better than previous comparable RNNs and LSTMs

Original MemNets I

- I (Input feature map)
  - pre-processing of data
- G (Generalization)
  - store input, compress input chunks, update previous inputs based on new ones
- O (Output feature map)
  - Calculates relevant memories to perform response (perhaps using score function)
- R (Response)
  - Formatting O into the desired output (e.g. a response sentence or word)

Original MemNets II

- Can use any ML algorithm for each component
- Components don’t necessarily all connect together
- Proposed MemNN instantiation is strongly supervised
    - k supporting sentences are chosen as input during training time

Overview of Experimental Evaluation I

- Increasing # of hops increases performance
- Embedding approaches improve performance
- Linear start improves performance
  - Remove softmax at memory layer during training, replace it when validation error stopped decreasing
- Huge improvement on tasks 17 and 19 with added non-linearity
  - higher embedding dimension (100 instead of 50)
  - ReLU after each hop

Overview of Experimental Evaluation II

	Baseline			MemN2N								
Task	Strongly Supervised MemNN [22]	LSTM [22]	MemNN WSH	BoW	PE	PE LS	PE LS RN	1 hop PE LS joint	2 hops PE LS joint	3 hops PE LS joint	PE LS RN joint	PE LS LW joint
1: 1 supporting fact	0.0	50.0	0.1	0.6	0.1	0.2	0.0	0.8	0.0	0.1	0.0	0.1
2: 2 supporting facts	0.0	80.0	42.8	17.6	21.6	12.8	8.3	62.0	15.6	14.0	11.4	18.8
3: 3 supporting facts	0.0	80.0	76.4	71.0	64.2	58.8	40.3	76.9	31.6	33.1	21.9	31.7
4: 2 argument relations	0.0	39.0	40.3	32.0	3.8	11.6	2.8	22.8	2.2	5.7	13.4	17.5
5: 3 argument relations	2.0	30.0	16.3	18.3	14.1	15.7	13.1	11.0	13.4	14.8	14.4	12.9
6: yes/no questions	0.0	52.0	51.0	8.7	7.9	8.7	7.6	7.2	2.3	3.3	2.8	2.0
7: counting	15.0	51.0	36.1	23.5	21.6	20.3	17.3	15.9	25.4	17.9	18.3	10.1
8: lists/sets	9.0	55.0	37.8	11.4	12.6	12.7	10.0	13.2	11.7	10.1	9.3	6.1
9: simple negation	0.0	36.0	35.9	21.1	23.3	17.0	13.2	5.1	2.0	3.1	1.9	1.5
10: indefinite knowledge	2.0	56.0	68.7	22.8	17.4	18.6	15.1	10.6	5.0	6.6	6.5	2.6
11: basic coreference	0.0	38.0	30.0	4.1	4.3	0.0	0.9	8.4	1.2	0.9	0.3	3.3
12: conjunction	0.0	26.0	10.1	0.3	0.3	0.1	0.2	0.4	0.0	0.3	0.1	0.0
13: compound coreference	0.0	6.0	19.7	10.5	9.9	0.3	0.4	6.3	0.2	1.4	0.2	0.5
14: time reasoning	1.0	73.0	18.3	1.3	1.8	2.0	1.7	36.9	8.1	8.2	6.9	2.0
15: basic deduction	0.0	79.0	64.8	24.3	0.0	0.0	0.0	46.4	0.5	0.0	0.0	1.8
16: basic induction	0.0	77.0	50.5	52.0	52.1	1.6	1.3	47.4	51.3	3.5	2.7	51.0
17: positional reasoning	35.0	49.0	50.9	45.4	50.1	49.0	51.0	44.4	41.2	44.5	40.4	42.6
18: size reasoning	5.0	48.0	51.3	48.1	13.6	10.1	11.1	9.6	10.3	9.2	9.4	9.2
19: path finding	64.0	92.0	100.0	89.7	87.4	85.6	82.8	90.7	89.9	90.2	88.0	90.6
20: agent's motivation	0.0	9.0	3.6	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2
Mean error (%)	6.7	51.3	40.2	25.1	20.3	16.3	13.9	25.8	15.6	13.3	12.4	15.2
Failed tasks (err. $> 5\%$)	4	20	18	15	13	12	11	17	11	11	11	10
On 10k training data												
Mean error (%)	3.2	36.4	39.2	15.4	9.4	7.2	6.6	24.5	10.9	7.9	7.5	11.0
Failed tasks (err. $> 5\%$)	2	16	17	9	6	4	4	16	7	6	6	6

[S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus, “End-to-End Memory Networks”, Nov 2015]

More Non-Linearity

	Baseline			MemN2N									
Task	Strongly Supervised MemNN	LSTM	MemNN WSH	BoW	PE	PE LS	PE LS RN	PE LS LW RN*	1 hop PE LS joint	2 hops PE LS joint	3 hops PE LS joint	PE LS RN joint	PE LS LW joint
1: 1 supporting fact	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2: 2 supporting facts	0.0	81.9	39.6	0.6	0.4	0.5	0.3	0.3	62.0	1.3	2.3	1.0	0.8
3: 3 supporting facts	0.0	83.1	79.5	17.8	12.6	15.0	9.3	2.1	80.0	15.8	14.0	6.8	18.3
4: 2 argument relations	0.0	0.2	36.6	31.8	0.0	0.0	0.0	0.0	21.4	0.0	0.0	0.0	0.0
5: 3 argument relations	0.3	1.2	21.1	14.2	0.8	0.6	0.8	0.8	8.7	7.2	7.5	6.1	0.8
6: yes/no questions	0.0	51.8	49.9	0.1	0.2	0.1	0.0	0.1	6.1	0.7	0.2	0.1	0.1
7: counting	3.3	24.9	35.1	10.7	5.7	3.2	3.7	2.0	14.8	10.5	6.1	6.6	8.4
8: lists/sets	1.0	34.1	42.7	1.4	2.4	2.2	0.8	0.9	8.9	4.7	4.0	2.7	1.4
9: simple negation	0.0	20.2	36.4	1.8	1.3	2.0	0.8	0.3	3.7	0.4	0.0	0.0	0.2
10: indefinite knowledge	0.0	30.1	76.0	1.9	1.7	3.3	2.4	0.0	10.3	0.6	0.4	0.5	0.0
11: basic coreference	0.0	10.3	25.3	0.0	0.0	0.0	0.0	0.1	8.3	0.0	0.0	0.0	0.4
12: conjunction	0.0	23.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
13: compound coreference	0.0	6.1	12.3	0.0	0.1	0.0	0.0	0.0	5.6	0.0	0.0	0.0	0.0
14: time reasoning	0.0	81.0	8.7	0.0	0.2	0.0	0.0	0.1	30.9	0.2	0.2	0.0	1.7
15: basic deduction	0.0	78.7	68.8	12.5	0.0	0.0	0.0	0.0	42.6	0.0	0.0	0.2	0.0
16: basic induction	0.0	51.9	50.9	50.9	48.6	0.1	0.4	51.8	47.3	46.4	0.4	0.2	49.2
17: positional reasoning	24.6	50.1	51.1	47.4	40.3	41.1	40.7	18,6	40.0	39.7	41.7	41.8	40.0
18: size reasoning	2.1	6.8	45.8	41.3	7.4	8.6	6.7	5.3	9.2	10.1	8.6	8.0	8.4
19: path finding	31.9	90.3	100.0	75.4	66.6	66.7	66.5	2.3	91.0	80.8	73.3	75.7	89.5
20: agent's motivation	0.0	2.1	4.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mean error (%)	3.2	36.4	39.2	15.4	9.4	7.2	6.6	4.2	24.5	10.9	7.9	7.5	11.0
Failed tasks (err. $> 5\%$)	2	16	17	9	6	4	4	3	16	7	6	6	6

[S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus, "End-to-End Memory Networks", Nov 2015]

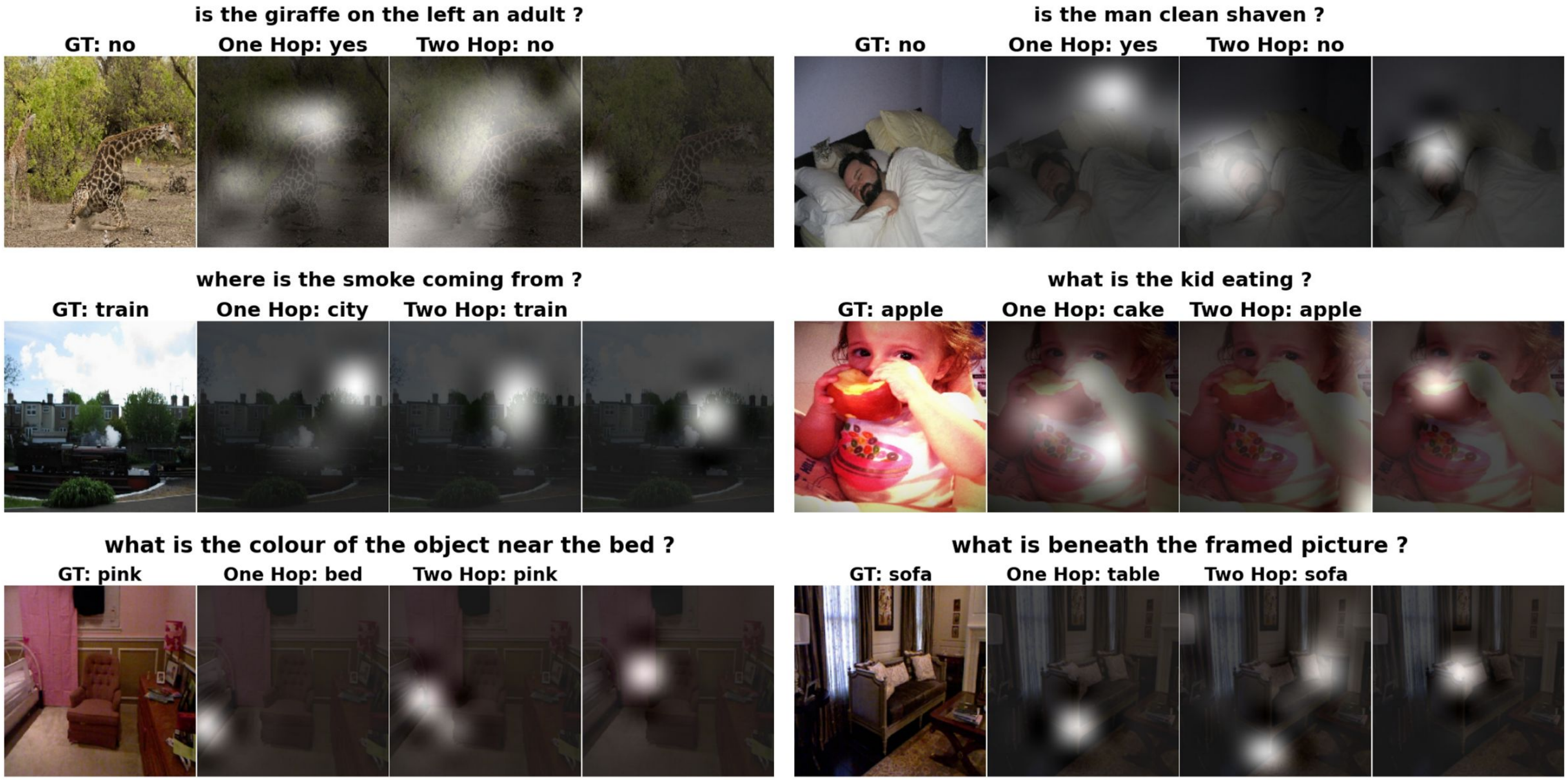
Strengths of MemN2N

- Less supervised than original MemNN
- Can be trained end-to-end
- Outperforms tuned RNNs and LSTMs for language modelling
- MemN2N - has ~1.5x params as vanilla RNN
  - LSTM is ~4x !

Weaknesses & Future Direction of MemN2N

- Another way to learn embeddings (skip-thought!)
- Multi-word answers
- Higher error than MemNN in a lot of tasks
- Handling more data in memory
  - what if it doesn't fit in memory? (databases)
  - what if it takes too long to look up? (bucketing)
- What about long stories?
- Positional reasoning could still use some work
  - 18.6% error at best

Applications



[H. Xu, K. Saenko, “Ask, Attend, and Answer: Exploring Question-Guided Spatial Attention for Visual Question-Answering”, Nov. 2015]

Movie	The Lord of the Rings: The Fellowship of the Ring	Harry Potter and the Chamber of Secrets	The Lord of the Rings: The Return of the King
Question	Who destroys Sauron in the battlefield?	What does Harry trick Lucius into doing?	Why does Arwen wish to stay in Middle Earth?
Video Shot			
Answer	Isildur	Freeing Dobby	Arwen sees her son with Aragorn in her visions
Option 1	Smeagol	Releasing Dobby to Harry's care	Because she is too weak to travel
Option 2	Gollum	Releasing Dobby to Dumbledore's care	Because she wants to die on Middle Earth
Option 3	The Ring	Releasing Dobby to Hagrid's care	Because she likes Middle Earth
Option 4	Bilbo	Admitting he put Tom Riddle's diary in Ginny's cauldron	Because her son asked to stay

[M. Tapaswi, Y. Zhu, et. el., "Movie QA: Understanding Stories in Movies Through Question-Answering", Dec. 2015]

DEMO

Source: https://github.com/vinhkhuc/MemN2N-babi-python

Questions?