Memory Network For Question Answering

Virgile Audi
Nicolas Drizard

TF: Sam Wiseman
May, 3^{rd} 2015

Outline

Table of Contents

Supporting Facts

Story

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.

Q Where is Mary? **A** office

Story

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.

Q Where was the apple before the kitchen? **A** office

Reasoning

Story

Sheep are afraid of wolves.
Cats are afraid of dogs.
Mice are afraid of cats.
Gertrude is a sheep.

Q What is Gertrude afraid of? **A** wolves

Story

Lily is a swan.
Lily is white.
Bernhard is green.
Greg is a swan.

Q What color is Greg? **A** white

Intuition

How do Humans build their answer?

1. Type of question
2. Occurence of the words from the question
3. Associations of words (*memory*)
4. Meaning of words (*reasoning*, *interpretation*)

Table of Contents

Count Based Model

Prediction based on two features:

$$\hat{y}(X, Q) = argmax(\log(f_1(X)) + \log(f_2(Q)))$$

with

- (X, Q) tuple story, question
- $f_1(X)$: answer words counts in the story (weighted by order of appearance)
- $f_2(Q)$: embedding of the question based on possible answers question word

Table of Contents

Single Hop Architecture

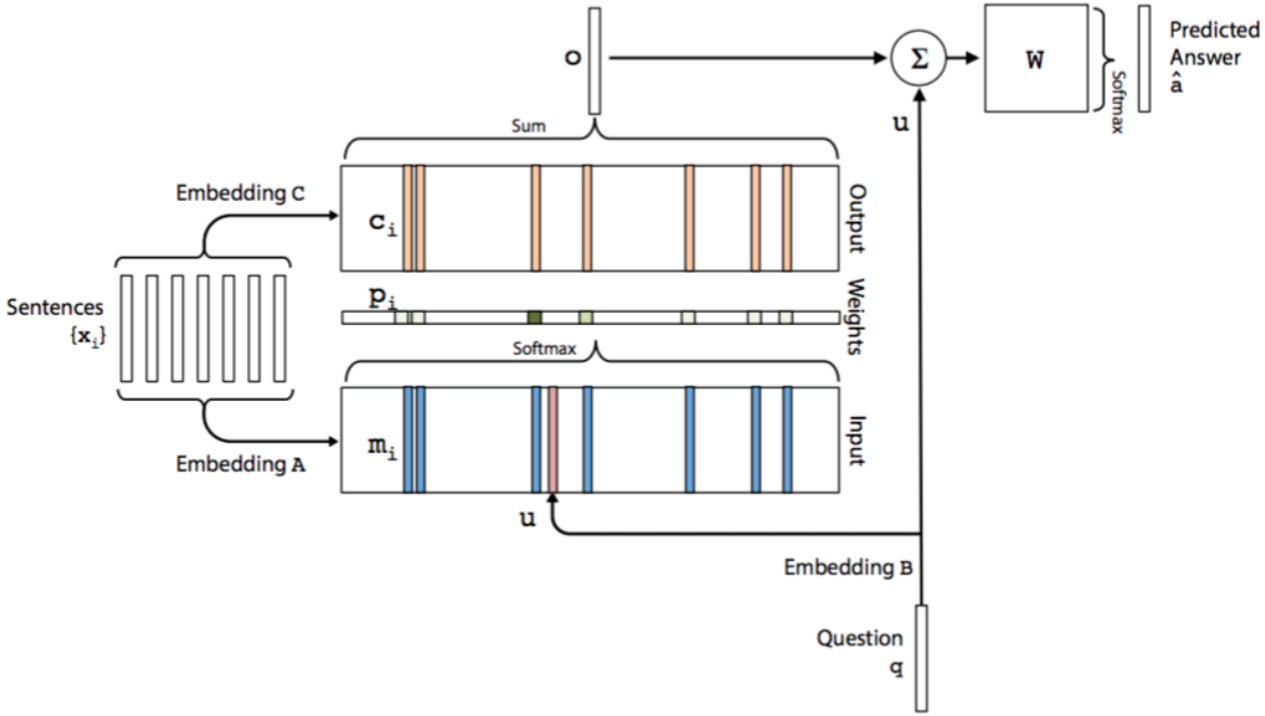

Figure: Single Hop architecture

Source: End-To-end Memory Networks

Multiple Hops Architecture

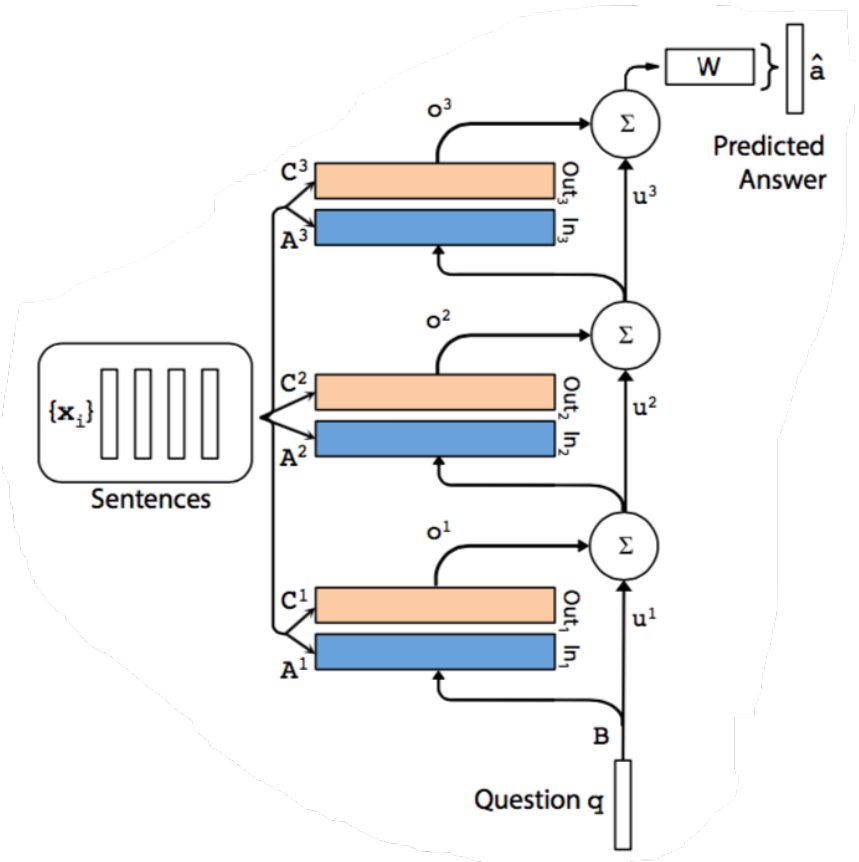

Figure: Multiple Hops architecture

Source: End-To-end Memory Networks

Parameters Tying

Two Approaches to reduce the number of parameters:

- **Adjacent**
  - $A^{k+1} = C^k$
  - $B = A^1$
- **RNN-like**
  - $A^1 = A^2 = ... = A^k$
  - $C^1 = C^2 = ... = C^k$
  - $u^{k+1} = Hu^k + o^k$

Implementation Tricks

- bag-of-words representation $x_i = \{x_{i1}, ..., x_{is}\}$ becomes $m_i = \sum_j A x_{ij}$
- Temporal encoding $m_i = \sum_j A x_{ij} + T_A(i)$
- high variance, best model over several training

Table of Contents

Accuracy Results on Train

Task	MemNN 1 hop adjacent	MemNN 3 hops adjacent	Baseline
1: 1 supporting fact	100	99	50
2: 2 supporting facts	16	90	40
3: 3 supporting facts	25	85	26
4: Two arg relation	71	79	34
5: Three arg relation	86	89	48
6: Yes/No	51	97	50
7: Counting	87	80	49
8: Lists/Sets	90	90	0
9: Simple Negation	64	86	64
10: Indefinite Knowledge	61	85	44
11: Basic coreference	80	87	63
12: Conjunction	100	100	60
13: Compound Coref	93	93	68
14: Time Reasoning	72	94	29
15: Basic Deduction	39	39	14
16: Basic Induction	50	54	52
17: Size Reasoning	53	62	48
18: Positional Reasoning	61	63	53
19: Path Finding	24	24	0
20: Agent's Motivation	92	92	52
TOTAL	**66**	**82**	**42.2**

Accuracy Results on

Task	MemNN 1 hop adjacent	MemNN 3 hops adjacent	Baseline
1: 1 supporting fact	100	99	50
2: 2 supporting facts	16	90	40
3: 3 supporting facts	25	85	26
4: Two arg relation	71	79	34
5: Three arg relation	86	89	48
6: Yes/No	51	97	50
7: Counting	87	80	49
8: Lists/Sets	90	90	0
9: Simple Negation	64	86	64
10: Indefinite Knowledge	61	85	44
11: Basic coreference	80	87	63
12: Conjunction	100	100	60
13: Compound Coref	93	93	68
14: Time Reasoning	72	94	29
15: Basic Deduction	39	39	14
16: Basic Induction	50	54	52
17: Size Reasoning	53	62	48
18: Positional Reasoning	61	63	53
19: Path Finding	24	24	0
20: Agent's Motivation	92	92	52
TOTAL	**66**	**82**	**42.2**

Task	MemNN 1 Hop	MemNN 3 Hop
1	99.7	99
2	17	83.9
3	19	65.2
4	66.4	66.9
5	82	81.5
6	49.6	92
7	85	78.2
8	89.3	NAN
9	63.5	82.3
10	57.4	83.2
11	83	88.3
12	99.9	100
13	94.4	94.4
14	65.6	84.8
15	23.1	23.3
16	46.9	46.6
17	47.5	54.3
18	51.6	52.4
19	15.8	NAN
20	90.8	0.91
TOTAL	62.36	75.9

Table of Contents

- Positional Encoding
- MC Test dataset
- Dynamic Memory