Memory Networks for Language Understanding

Antoine Bordes - Facebook AI Research

LXMLS – Lisbon July 28, 2016

Bots?


AT&T 4G
7:40 AM
What can I help you with?
“ Play a good song ”
Sorry, I couldn’t find ‘a good song’ in your music.

End-to-End Dialog Agents

We believe a true dialog agent should:

- Be able to **combine all its knowledge and reason** to fulfill complex tasks
- Handle **long open-ended conversations** involving effectively **tracking and predicting dialog and world states**
- Be able to **learn** (new tasks) and **acquire knowledge** via conversation and reading (and observing the world in multimodal scenarios).

Our directions:

1. **Machine Learning End-to-End systems**

Memory Networks (Weston et al., ICLR15; Sukhbaatar et al., NIPS15)

- Class of models that combine large memory with learning component that can read and write to it.
- Incorporates **reasoning** with **attention** over **memory** (RAM).
- Most ML has limited memory which is more-or-less all that's needed for "low level" tasks e.g. object detection.

Our motivation: long-term memory is required to read a story and then e.g. answer questions about it.

Similarly, it's also required for **dialog**: to remember previous dialog (short- and long-term), and respond.

bAbI Tasks (Weston et al., ICLR16)

- Set of 20 tasks testing basic reasoning capabilities for QA from stories
- Short stories are generated from a simulation
- Easy to interpret results / test a broad range of properties

John dropped the milk.
John took the milk there.
Sandra went to the bathroom.
John moved to the hallway.
Mary went to the bedroom.
Where is the milk ? Hallway

Task 3: Two supporting facts

The suitcase is bigger than the chest.
The box is bigger than the chocolate.
The chest is bigger than the chocolate.
The chest fits inside the container.
The chest fits inside the box.
Does the suitcase fit in the chocolate?
no

Task 18: Size reasoning

- Useful to foster innovation: cited ~100 times

Example

Simple grammar

Command format

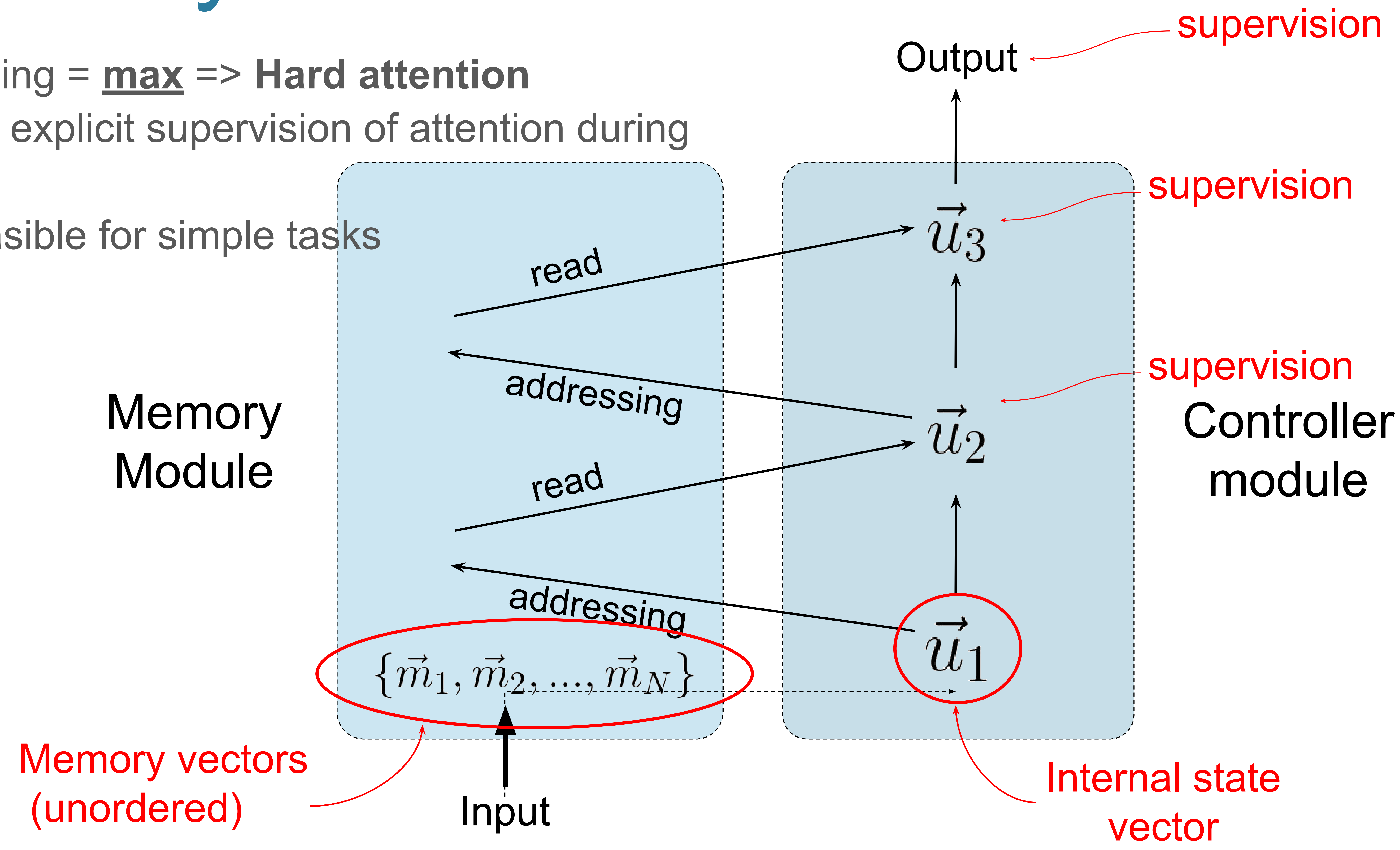
```
jason go kitchen
jason get milk
jason go office
jason drop milk
jason go bathroom
where is milk ?   A: office
where is jason? A: bathroom
```

Story

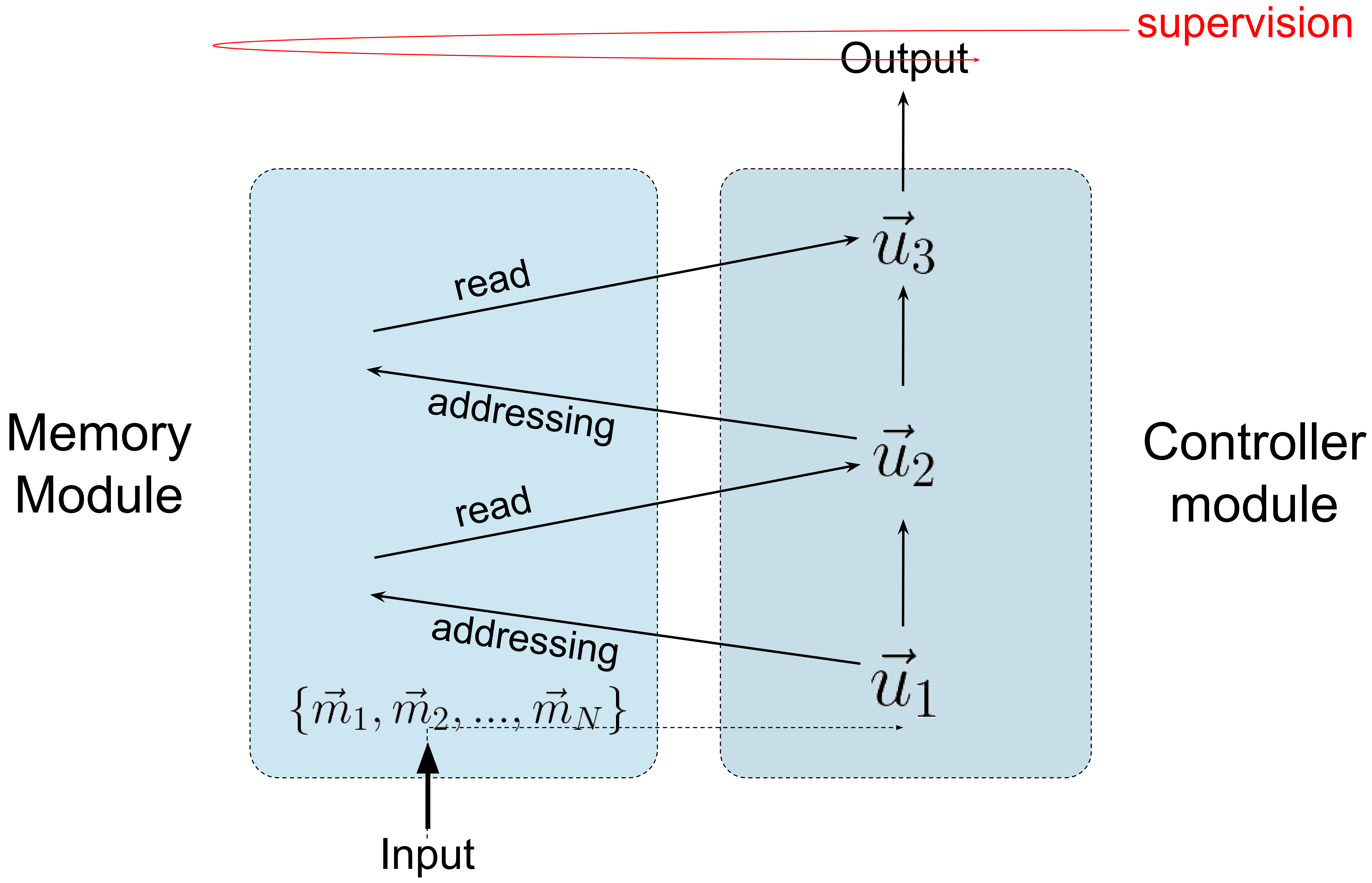
Jason went to the kitchen.
Jason picked up the milk.
Jason travelled to the office.
Jason left the milk there.
Jason went to the bathroom.
Where is the milk now? A: office
Where is Jason? A: bathroom

Memory Networks (Weston et al., ICLR15)

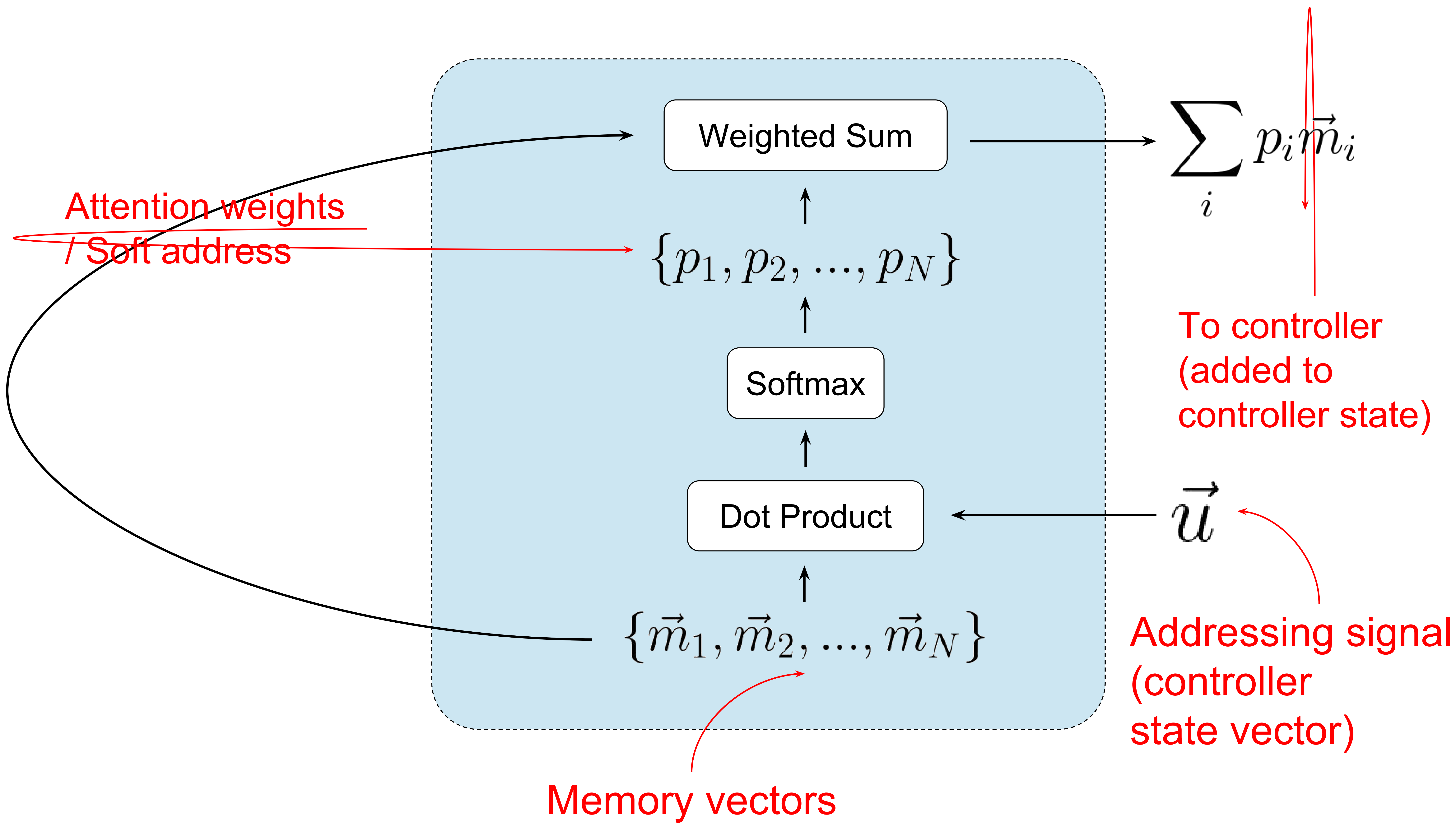
- Addressing = **<u>max</u>** => **Hard attention**
- requires explicit supervision of attention during training
- Only feasible for simple tasks



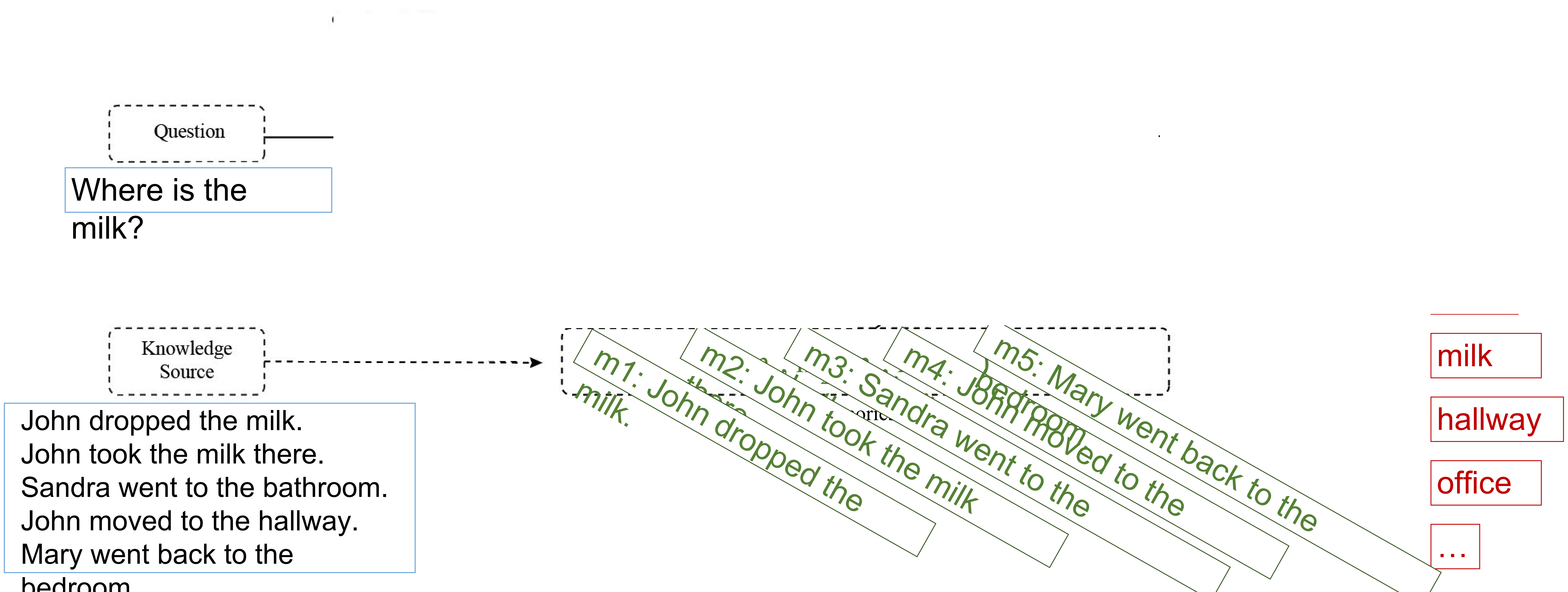
End-to-end Memory Networks (Sukhbaatar et al., NIPS15)



Memory Module



Memory Networks on bAbI



Dashboard

Training on 1k stories

TASK	Weakly supervised			Supervised Supp. Facts	
	N-grams	LSTMs	MemN2N	Memory Networks	StructSVM+ coref+srl
T1. Single supporting fact	36	50	PASS	PASS	PASS
T2. Two supporting facts	2	20	87	PASS	74
T3. Three supporting facts	7	20	60	PASS	17
T4. Two arguments relations	50	61	PASS	PASS	PASS
T5. Three arguments relations	20	70	87	PASS	83
T6. Yes/no questions	49	48	92	PASS	PASS
T7. Counting	52	49	83	85	69
T8. Sets	40	45	90	91	70
T9. Simple negation	62	64	87	PASS	PASS
T10. Indefinite knowledge	45	44	85	PASS	PASS
T11. Basic coreference	29	72	PASS	PASS	PASS
T12. Conjunction	9	74	PASS	PASS	PASS
T13. Compound coreference	26	PASS	PASS	PASS	PASS
T14. Time reasoning	19	27	PASS	PASS	PASS
T15. Basic deduction	20	21	PASS	PASS	PASS
T16. Basic induction	43	23	PASS	PASS	24
T17. Positional reasoning	46	51	49	65	61
T18. Size reasoning	52	52	89	PASS	62
T19. Path finding	0	8	7	36	49
T20. Agent's motivation	76	91	PASS	PASS	PASS

Attention during memory lookups

Samples from toy QA tasks

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				

20 bAbI Tasks

	Test Acc	Failed tasks
MemNN	93.3%	4
LSTM	49%	20
MemN2N 1 hop	74.82%	17
2 hops	84.4%	11
3 hops	87.6.%	11

Related Memory Models

(published before or ~same time as original paper)

- RNNSearch (Bahdanau et al.) for Machine Translation
  - Can be seen as a Memory Network where memory goes back only one sentence (writes embedding for each word).
  - At prediction time, reads memory and performs a soft max to find best alignment (most useful words). 1 hop only.
- Generating Sequences With RNNs (Graves, '13)
  - Also does alignment with previous sentence to generate handwriting (so RNN knows what letter it's currently on).
- Neural Turing Machines (Graves et al., 14) [on arxiv just 5 days after MemNNs!]
  - Has read and write operations over memory to perform tasks (e.g. copy, sort, associative recall).
  - 128 memory slots in experiments; content addressing computes a score for each slot → slow for large memory?
- Earlier work by (Das '92), (Schmidhuber et al., 93), DISCERN (Miikkulainen, '90) and others...

Learning of Basic Algorithms using Reasoning, Attention, Memory (RAM)
(e.g. addition, multiplication, sorting)

Methods include adding stacks and addressable memory to RNNs:

- “Neural Net Architectures for Temporal Sequence Processing” M. Mozer.
- “Neural Turing Machines” A. Graves, G. Wayne, I. Danihelka.
- “Inferring Algorithmic Patterns with Stack Augmented Recurrent Nets” A. Joulin, T. Mikolov
- “Learning to Transduce with Unbounded Memory” E. Grefenstette et al.
- “Neural Programmer-Interpreters” S. Reed, N. de Freitas.
- “Reinforcement Learning Turing Machine” W. Zaremba and I. Sutskever.
- “Learning Simple Algorithms from Examples” W. Zaremba, T. Mikolov, A. Joulin, R. Fergus
- “The Neural GPU and the Neural RAM machine” I. Sutskever.

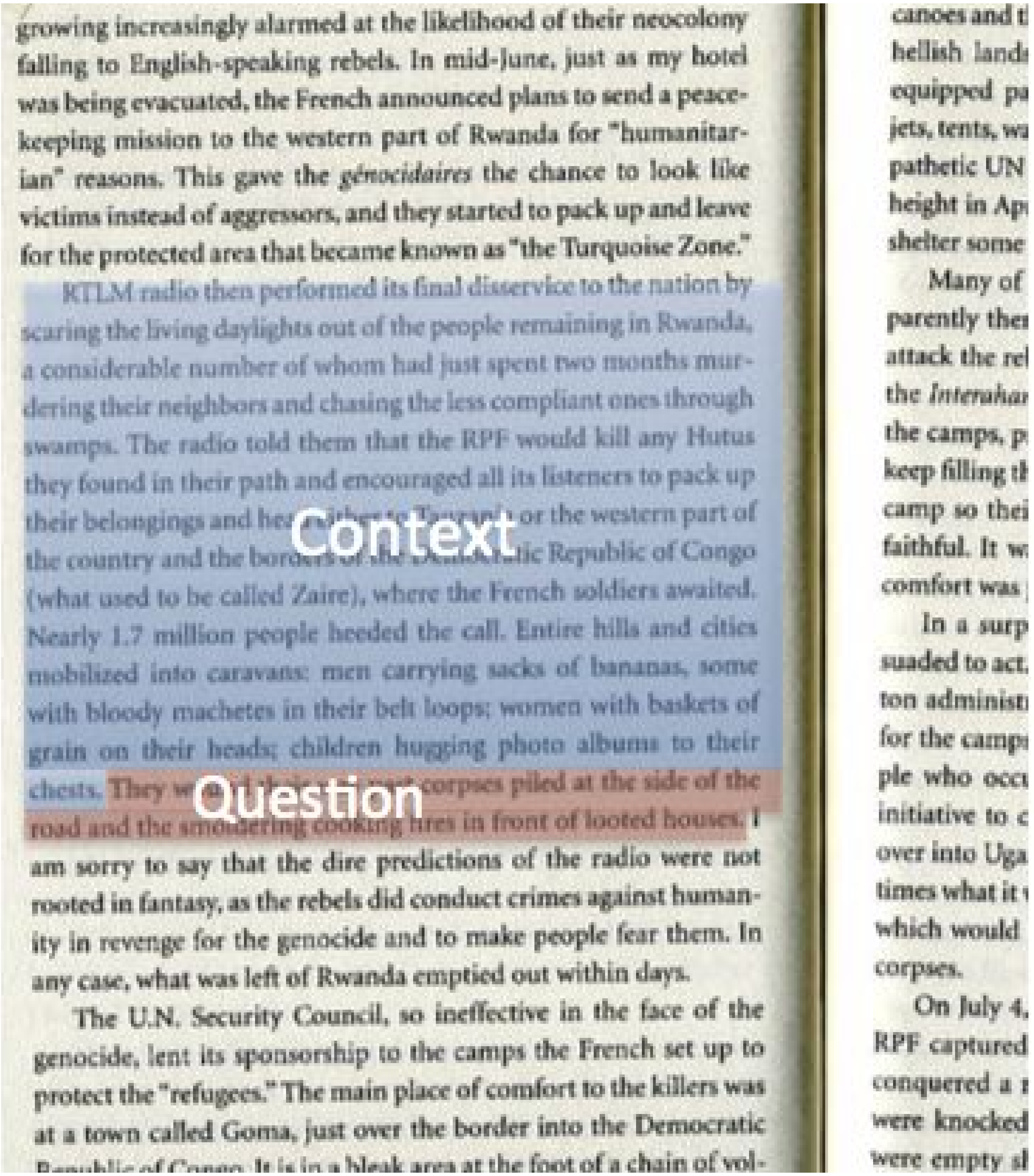
How about on real data? In other conditions?

- Toy AI tasks are important for developing innovative methods.
- But they do not give all the answers.

- How do these models work in real/different conditions?
  - Story understanding (Children's Book, News articles)
  - Open Question Answering (Knowledge Bases, Wikipedia)
  - Dialog (Synthetic Dialog, Ubuntu)

Story Understanding

Children's Books Test (CBT) (Hill et al., ICLR16)



growing increasingly alarmed at the likelihood of their neocolony falling to English-speaking rebels. In mid-June, just as my hotel was being evacuated, the French announced plans to send a peace-keeping mission to the western part of Rwanda for "humanitarian" reasons. This gave the *génocidaires* the chance to look like victims instead of aggressors, and they started to pack up and leave for the protected area that became known as "the Turquoise Zone."

KTLM radio then performed its final disservice to the nation by scaring the living daylights out of the people remaining in Rwanda, a considerable number of whom had just spent two months murdering their neighbors and chasing the less compliant ones through swamps. The radio told them that the RPF would kill any Hutus they found in their path and encouraged all its listeners to pack up their belongings and he[...] Turquoise" or the western part of the country and the borders of the [...]ic Republic of Congo (what used to be called Zaire), where the French soldiers awaited. Nearly 1.7 million people heeded the call. Entire hills and cities mobilized into caravans: men carrying sacks of bananas, some with bloody machetes in their belt loops; women with baskets of grain on their heads; children hugging photo albums to their chests. They w[...] corpses piled at the side of the road and the smoldering cooking fires in front of looted houses. I am sorry to say that the dire predictions of the radio were not rooted in fantasy, as the rebels did conduct crimes against humanity in revenge for the genocide and to make people fear them. In any case, what was left of Rwanda emptied out within days.

The U.N. Security Council, so ineffective in the face of the genocide, lent its sponsorship to the camps the French set up to protect the "refugees." The main place of comfort to the killers was at a town called Goma, just over the border into the Democratic Republic of Congo. It is in a bleak area at the foot of a chain of vol-

Story understanding dataset based on 118 children books from project Gutenberg

1 `` phebe beckoned to him ; i saw her , '' cried rose , staring hard at the door .
2 `` is it more presents coming ? ''
3 asked jamie , just as his brother re-appeared , looking more excited than ever .
4 `` yes ; a present for mother , and here it is ! ''
5 roared archie , flinging wide the door to let in a tall man , who cried out , `` where 's my little woman ?
6 the first kiss for her , then the rest may come on as fast as they like . ''
7 before the words were out of his mouth , mrs. jessie was half-hidden under his rough great-coat , and four boys were prancing about him clamouring for their turn .
8 of course , there was a joyful tumult for a time , during which rose slipped into the window recess and watched what went on , as if it were a chapter in a christmas story .
9 it was good to see bluff uncle jem look proudly at his tall son , and fondly hug the little ones .
10 it was better still to see him shake his brothers ' hands as if he would never leave off , and kiss all the sisters in a way that made even solemn aunt myra brighten up for a minute .

11 but it was best of all to see him finally established in grandfather 's chair , with his `` little woman '' beside him , his three youngest boys in his lap , and _____ hovering over him like a large-sized cherub .

faith | brothers | rose | <u>**archie**</u> | rest | mouth | way | mother | sisters | george

Memory Networks on CBT

Memories format?

- Sentence: whole sentences
  (as in the bAbI tasks)
- Word: 1 word at a time
  (language modeling style)
- Words window: store windows made through the story
  (convolution style)

S: 1 So they had to fall a long way .
2 So they got their tails fast in their mouths .
3 So they could n't get them out again .
4 That 's all . '
5 ` Thank you , ' said Alice , ` it 's very interesting .
6 I never knew so much about a whiting before . ' '
7 I can tell you more than that , if you like , ' said the Gryphon .
8 ` Do you know why it 's called a whiting ? ''
9 I never thought about it , ' said Alice .
10 ` Why ? '
11 ` IT DOES THE BOOTS AND SHOES . '
12 the Gryphon replied very solemnly .
13 Alice was thoroughly puzzled .
14 ` Does the boots and shoes ! '
15 she repeated in a wondering tone .
16 ` Why , what are YOUR shoes done with ? '
17 said the Gryphon . '
18 I mean , what makes them so shiny ? '
19 Alice looked down at them , and considered a little before she gave her answer .
20 ` They 're done with blacking , I believe .

q: `Boots and shoes under the sea , ' the _________ went on in a deep voice , are done with a whiting .

C: Alice, BOOTS, Gryphon, SHOES, answer, fall, mouths, tone, way, whiting.

Different Word Types / Different Models

Methods	Named Entities	Common Nouns	Verbs	Prepositions
Humans (query)(*)	0.520	0.644	0.716	0.676
Humans (context+query)(*)	***0.816***	***0.816***	***0.828***	0.708
Maximum frequency (corpus)	0.120	0.158	0.373	0.315
Maximum frequency (context)	0.335	0.281	0.285	0.275
Sliding window	0.168	0.196	0.182	0.101
Word distance model	0.398	0.364	0.380	0.237
Kneser-Ney language model	0.390	0.544	0.778	0.768
Kneser-Ney language model + cache	0.439	0.577	0.772	0.679
Embedding Model (context+query)	0.253	0.259	0.421	0.315
Embedding Model (query)	0.351	0.400	0.614	0.535
Embedding Model (window)	0.362	0.415	0.637	0.589
Embedding Model (window+position)	0.402	0.506	0.736	0.670
LSTMs (query)	0.408	0.541	0.813	0.802
LSTMs (context+query)	0.418	0.560	**0.818**	0.791
Contextual LSTMs (window context)	0.436	0.582	0.805	**0.806**
MemNNs (lexical memory)	0.431	0.562	0.798	0.764
MemNNs (window memory)	0.493	0.554	0.692	0.674
MemNNs (sentential memory + PE)	0.318	0.305	0.502	0.326
MemNNs (window memory + self-sup.)	**0.666**	**0.630**	0.690	0.703

Question Answering on New's Articles

We evaluate our models on the data from:

"Teaching Machines to Read and Comprehend"

Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will

METHODS	VALIDATION	TEST
MAXIMUM FREQUENCY (ARTICLE)$^{(*)}$	0.305	0.332
SLIDING WINDOW	0.005	0.006
WORD DISTANCE MODEL$^{(*)}$	0.505	0.509
DEEP LSTMS (ARTICLE+QUERY)$^{(*)}$	0.550	0.570
CONTEXTUAL LSTMS ("ATTENTIVE READER")$^{(*)}$	0.616	0.630
CONTEXTUAL LSTMS ("IMPATIENT READER")$^{(*)}$	0.618	0.638
MEMNNS (WINDOW MEMORY)	0.580	0.606
MEMNNS (WINDOW MEMORY + SELF-SUP.)	0.634	0.668
MEMNNS (WINDOW MEMORY + ENSEMBLE)	0.612	0.638
MEMNNS (WINDOW MEMORY + SELF-SUP. + ENSEMBLE)	0.649	0.684
MEMNNS (WINDOW + SELF-SUP. + ENSEMBLE + EXCLUD. COOCURRENCES)	**0.662**	**0.694**

Table 3: **Results on CNN QA.** $^{(*)}$Results taken from Hermann et al. (2015).

Answer	
Oisin Tymon	*ent193*

Latest *Fresh* Results

- Our best results:

  QACNN: 69.4 CBT-NE: 66.6 CBT-CN: 63.0

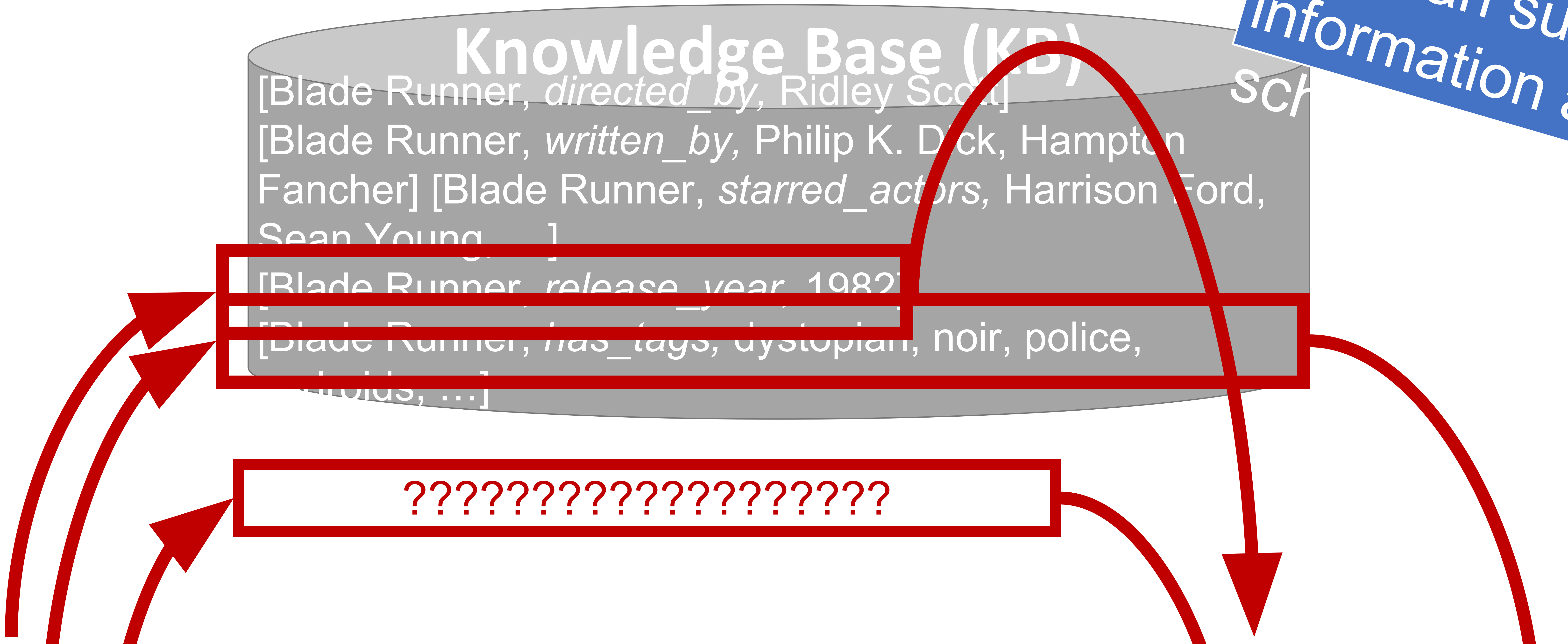
- Text Understanding with the Attention Sum Reader Network. *Kadlec et al.* **(4 Mar ‘16)** QACNN: 75.4 CBT-NE: 71.0 CBT-CN: 68.9
- Iterative Alternating Neural Attention for Machine Reading. *Sordoni et al.* **(7 Jun ’16)** QACNN: 76.1 CBT-NE: 72.0 CBT-CN: 71.0
- Natural Language Comprehension with the EpiReader. *Trischler et al.* **(7 Jun ’16)** QACNN: 74.0 CBT-NE: 71.8 CBT-CN: 70.6
- Gated-Attention Readers for Text Comprehension. *Dhingra et al.* **(5 Jun ’16)** QACNN: 77.4 CBT-NE: 71.9 CBT-CN: 69.0

Open Question Answering

Open-domain Question Answering

Answer questions on any topic

KBs can suffer from missing information and fixed



What year was the movie Blade Runner released?	1982
Can you describe Blade Runner in a few words?	A dystopian and noir movie
In Blade Runner, who built the Replicants?	???

Information Extraction

Wikipedia Entry: Blade Runner

Blade Runner is a 1982 American neo-noir dystopian science fiction film directed by Ridley Scott and starring Harrison Ford, Rutger Hauer, Sean Young, and Edward James Olmos. The screenplay, written by Hampton Fancher and David Peoples, is a modified film adaptation of the 1968 novel “Do Androids Dream of Electric Sheep?” by Philip K. Dick. The film depicts a dystopian Los Angeles in November 2019 in which genetically engineered replicants, which are visually indistinguishable from adult humans, are manufactured by the powerful Tyrell Corporation as well as by other “mega-corporations” around the world…

Knowledge Base (KB)

[Blade Runner, *directed_by,* Ridley Scott]
[Blade Runner, *written_by,* Philip K. Dick, Hampton Fancher] [Blade Runner, *starred_actors,* Harrison Ford, Sean Young, …]
[Blade Runner, *release_year,* 1982]
[Blade Runner, *has_tags,* dystopian, noir, police,
[Replicants, *manufactured_by,* Tyrell Corporation]

IE is not an easy

What year was the movie Blade Runner released?	1982 ?
Can you describe Blade Runner in a few words?	A dystopian and noir movie
In Blade Runner, who built the Replicants?	??? Tyrell Corporation

Question Answering Directly from Text

Wikipedia Entry: Blade Runner

Blade Runner is a 1982 American neo-noir dystopian science fiction film directed by Ridley Scott and starring Harrison Ford, Rutger Hauer, Sean Young, and Edward James Olmos. The screenplay, written by Hampton Fancher and David Peoples, is a modified film adaptation of the 1968 novel “Do Androids Dream of Electric Sheep?” by Philip K. Dick. The film depicts a dystopian Los Angeles in November 2019 in which genetically engineered replicants, which are visually indistinguishable from adult humans, are manufactured by the powerful Tyrell Corporation as well as by other “mega-corporations” around the world…

Much more information than in

But QA is

What year was the movie Blade Runner released? — 1982

Can you describe Blade Runner in a few words? — A dystopian and noir movie

In Blade Runner, who built the Replicants? — Tyrell Corporation

MovieQA (Miller et al., arxiv16)

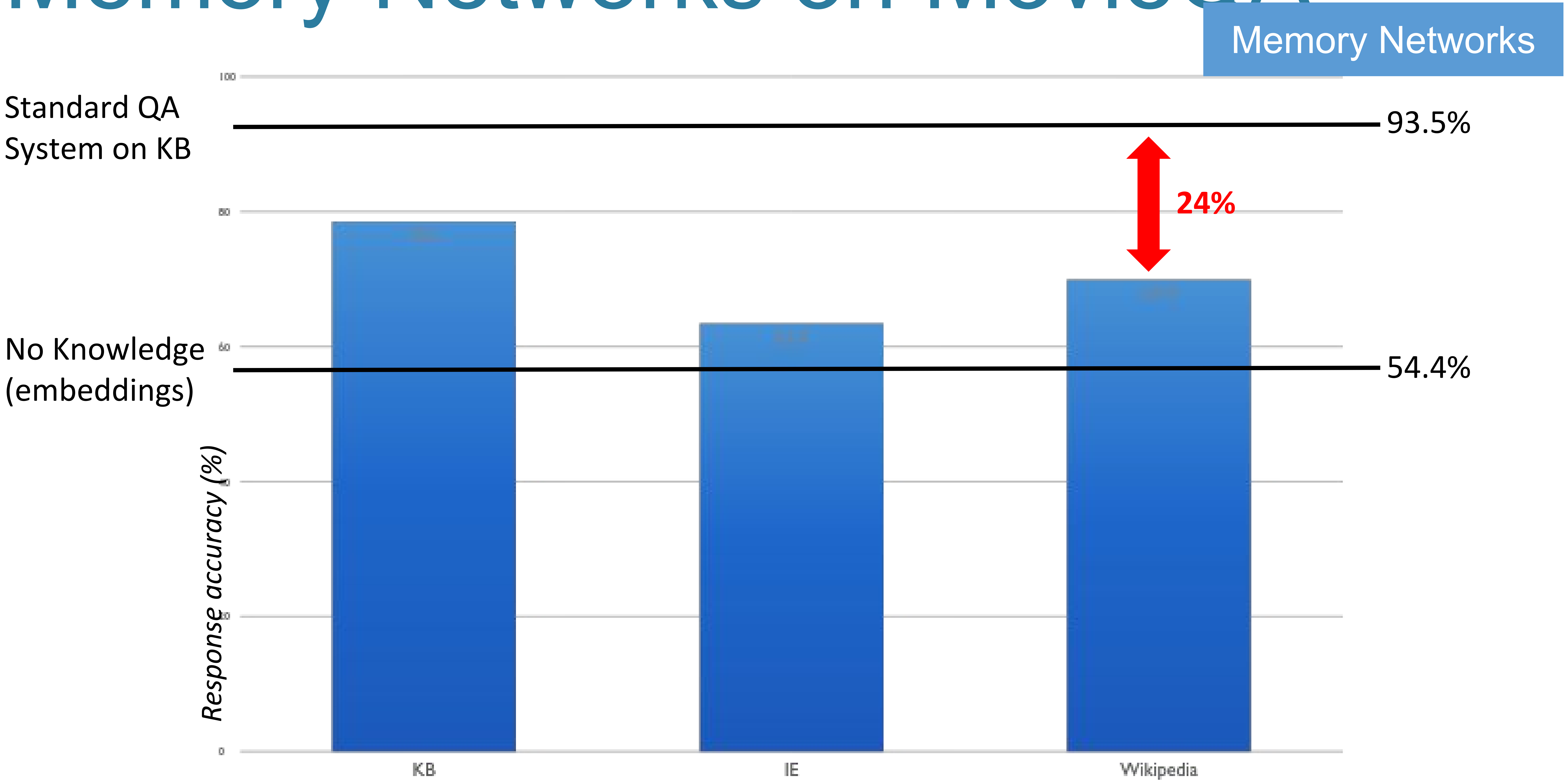
- Hypothesis: Systems answering from text directly must be on par with systems using KBs for questions whose answers are in KBs.

- MovieQA: a new analysis tool for QA
  - A set of 100k question -- answer pairs (based on SimpleQuestions)
  - 3 knowledge sources:
    - A KB based on OMDb
    - Raw text extracted from Wikipedia
    - An imperfect KB made by an IE system ran on the Wikipedia articles
  - Answers to all questions are in the KB and in the Wikipedia text.

Memory Networks for QA from KB (Bordes et al., arxiv15)

Memory Networks for QA from Text (Hill et al., ICLR16)

Memory Networks on MovieQA



Structuring Memories

- Structure in the symbolic memories
  - Parts of the memories match questions where others encode response

[Blade Runner, *directed_by*, Ridley Scott]

directed by Ridley Scott and starring

[Blade Runner, *release_year*, 1982]

is a 1982 American neo-noir

- Prior knowledge on the task
  - Which Wikipedia page do the windows come from?
  - Which knowledge source do memories have been extracted from?

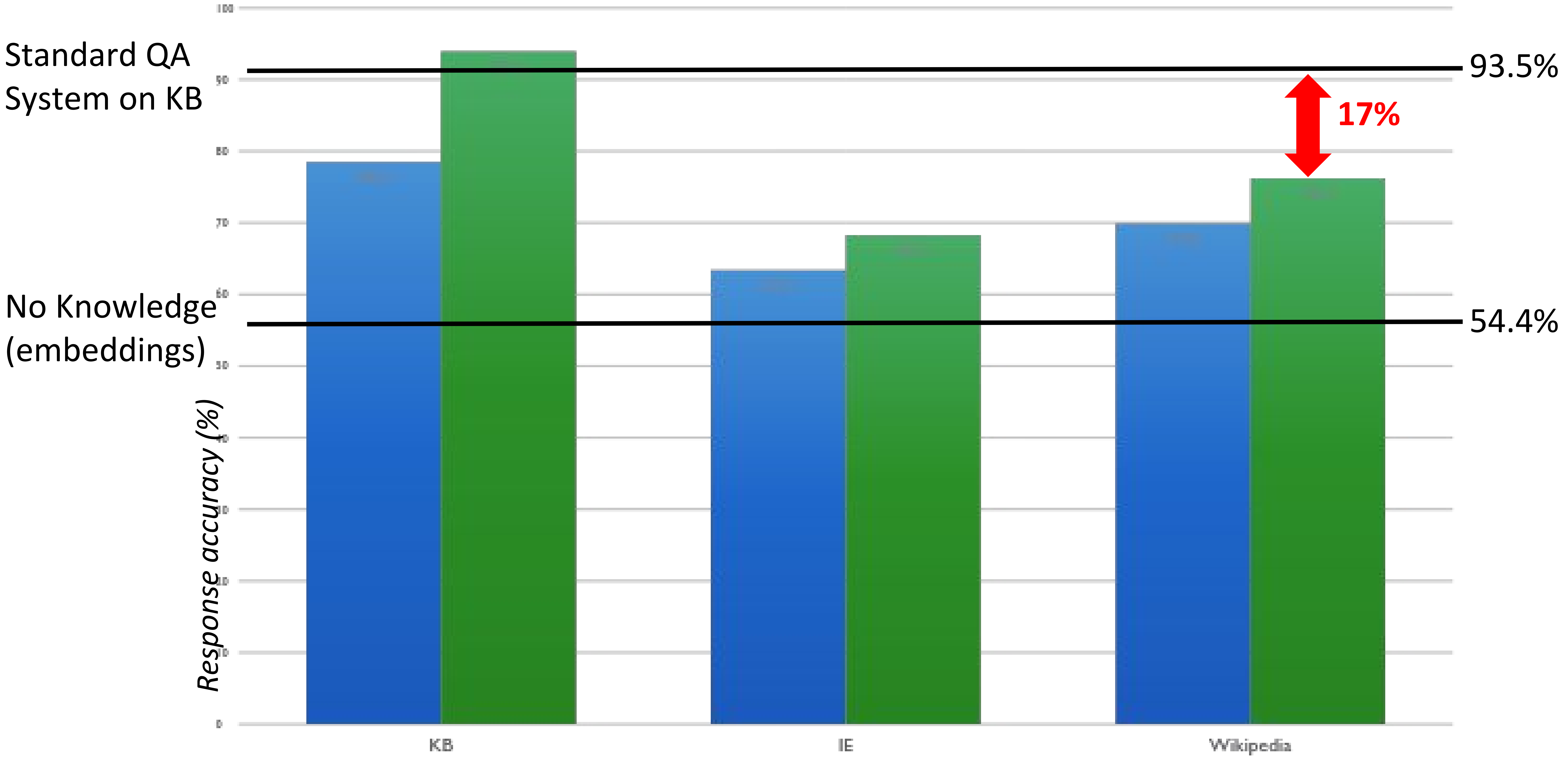
Key-Value Memory Networks on KB

Key-Value Memory Networks on Text

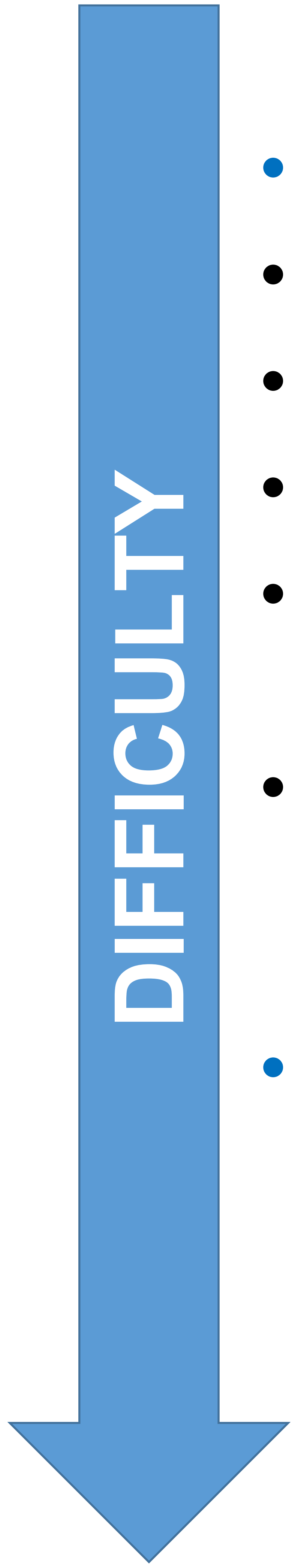
Results on MovieQA

Memory Networks

Key-Value Memory Networks

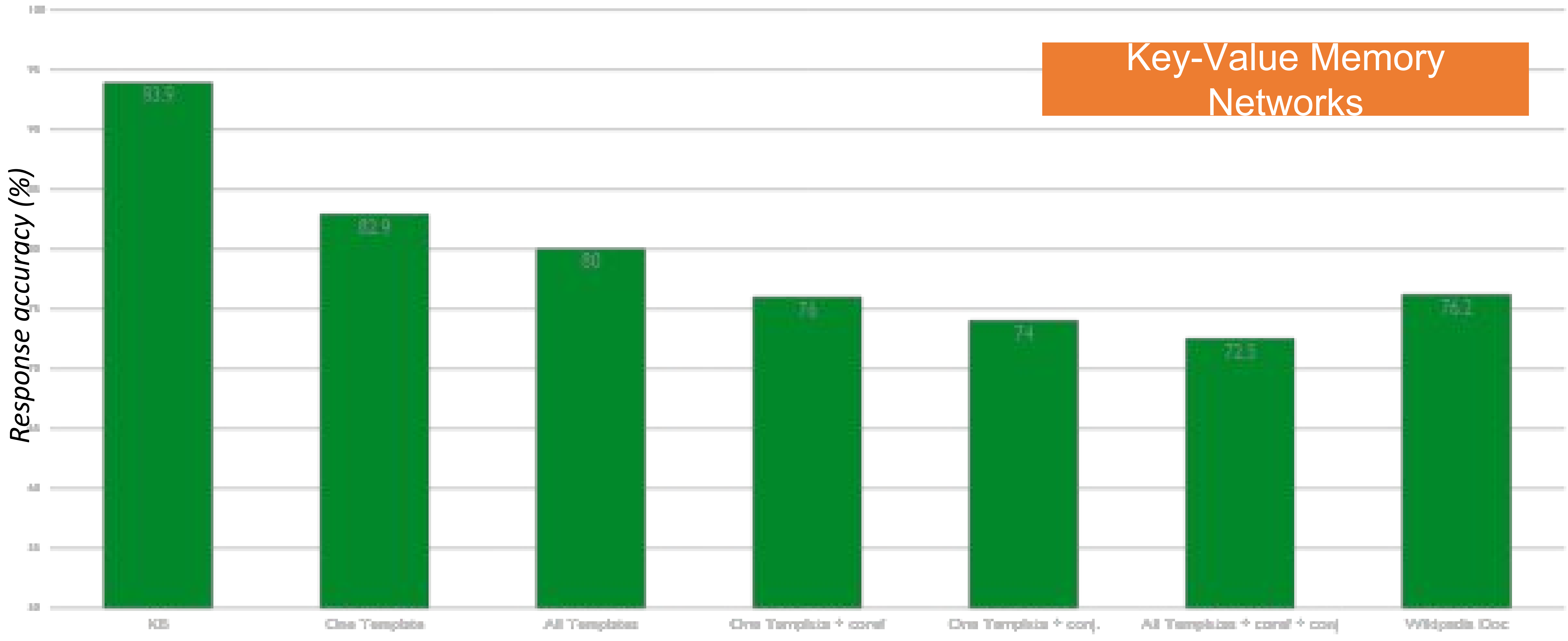


Synthetic Documents



- KB: [Flags of Our Fathers, *directed_by*, Clint Eastwood]
- One Template: *Clint Eastwood directed Flags of Our Fathers*
- All Templates: *Flags of Our Fathers was directed by Clint Eastwood.*
- One Template + coref.: *Flags of Our Fathers came out in 2006. Clint Eastwood directed it.*
- One Template + conjunctions: *Flags of Our Fathers is in English and Clint Eastwood directed Flags of Our Fathers.*
- All Templates + coref. + conj.: *Flags of Our Fathers is a famous film. Ryan Phillippe, Jesse Bradford, Adam Beach, and John Benjamin Hickey are the actors in it and Clint Eastwood is the person who directed it.*
- Wikipedia: *The film adaptation Flags of Our Fathers, which opened in the U.S. on October 20, 2006, was directed by Clint Eastwood and produced by Steven Spielberg, with a screenplay written by William Broyles, Jr. and Paul Haggis.*

Synthetic Documents Analysis



WikiQA (Yang et al., EMNLP15)

- QA Benchmark in the answer selection setting
- Key-Value Memories -> (window, sentence)
  - Q: How are glacier caves
  - A: A glacier cave is a cave
- Training size is very smal
  - Word embeddings pre-tr
  - Dropout regularization

Method	MAP	MRR
Word Cnt	0.4891	0.4924
Wgt Word Cnt	0.5099	0.5132
2-gram CNN (Yang *et al.*, 2015)	0.6520	0.6652
AP-CNN (Santos *et al.*, 2016)	0.6886	0.6957
Attentive LSTM (Miao *et al.*, 2015)	0.6886	0.7069
Attentive CNN (Yin and Schütze, 2015)	0.6921	0.7108
L.D.C. (Wang *et al.*, 2016)	0.7058	0.7226
Memory Network	0.5170	0.5236
Key-Value Memory Network	**0.7069**	**0.7265**

Dialog

How about on dialog data?

- Everything we showed so far was Q&A potentially with long-term context.
- We have also built a Movie Dialog Dataset (Dodge et al., ICLR16)

  Closed, but large, domain about movies (75k entities, 3.5M ex).
  - Ask facts about movies?
  - Ask for opinions (recommendations) about movies?
  - Dialog combining facts and opinions?
  - General chit-chat about movies (statements not questions)?

And… combination of all above in one end-to-end model.

Combines QA with Dialog Tasks (Dodge et al., ICLR16)

(Dialog 1) QA: *facts about movies*

Sample input contexts and target replies (in red) from Dialog Task 1:

What movies are about open source? Revolution OS
Ruggero Raimondi appears in which movies? Carmen
What movies did Darren McGavin star in? Billy Madison, The Night Stalker, Mrs. Pollifax-Spy, The Challenge
Can you name a film directed by Stuart Ortiz? Grave Encounters
Who directed the film White Elephant? Pablo Trapero
What is the genre of the film Dial M for Murder? Thriller, Crime
What language is Whity in? German

(Dialog 2) Recs: *movie recommendations*

Sample input contexts and target replies (in red) from Dialog Task 2:

Schindler's List, The Fugitive, Apocalypse Now, Pulp Fiction, and The Godfather are films I really liked. Can you suggest a film?
The Hunt for Red October

Some movies I like are Heat, Kids, Fight Club, Shaun of the Dead, The Avengers, Skyfall, and Jurassic Park. Can you suggest something else I might like? Ocean's Eleven

(Dialog 3) QA+Recs: *combination dialog*

Sample input contexts and target replies (in red) from Dialog Task 3:

I loved Billy Madison, Blades of Glory, Bio-Dome, Clue, and Happy Gilmore. I'm looking for a Music movie. School of Rock
What else is that about? Music, Musical, Jack Black, school, teacher, Richard Linklater, rock, guitar
I like rock and roll movies more. Do you know anything else?
Little Richard

(Dialog 4) Reddit: *real dialog*

Sample input contexts and target replies (in red) from Dialog Task 4:

I think the Terminator movies really suck, I mean the first one was kinda ok, but after that they got really cheesy. Even the second one which people somehow think is great. And after that... forgeddabotit.
C'mon the second one was still pretty cool.. Arny was still so badass, as was Sararah Connor's character.. and the way they blended real action and effects was perhaps the last of its kind...

Ubuntu Data (Lowe et al. McGill, ‘15)

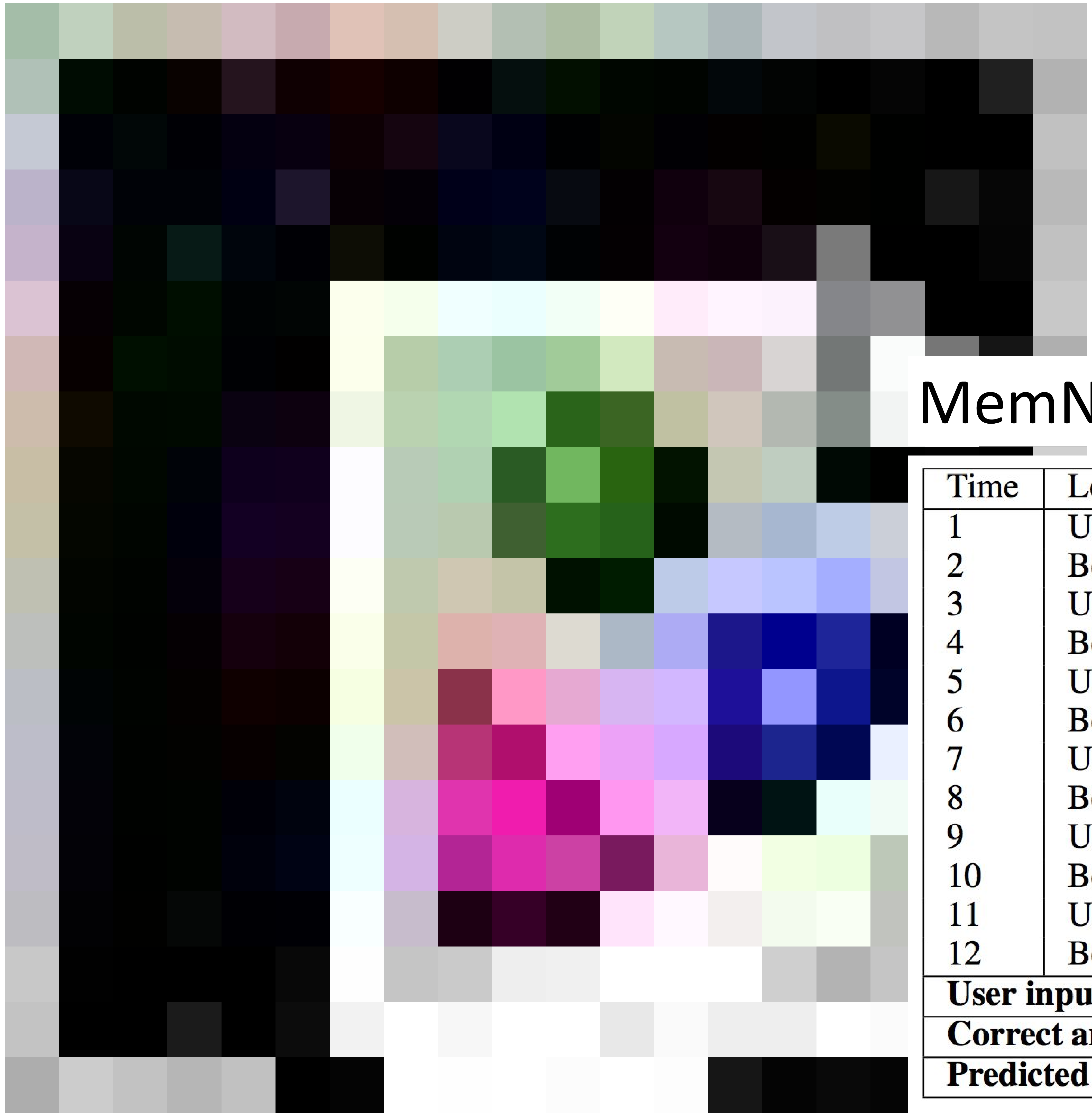
Dialog dataset: Ubuntu IRC channel logs, users ask questions about issues they are having with Ubuntu and get ans

METHODS		VALIDATION (HITS@1)	TEST (HITS@1)
IR†		N/A	48.81
RNN†		N/A	37.91
LSTM†		N/A	55.22
MEMN2N	1-HOP	57.23	56.25
MEMN2N	2-HOPS	64.28	63.51
MEMN2N	3-HOPS	64.31	63.72
MEMN2N	4-HOPS	64.01	62.82

Table 7: **Ubuntu Dialog Corpus results.** The evaluation is retrieval-based, similar to that of Reddit (Task 4). For each dialog, the correct answer is mixed among 10 random candidates; Hits@1 (in %) are reported. Methods with † have been ran by Lowe et al. (2015).

Best current results: “Sentence Pair Scoring: Towards Unified Framework for Text Comprehension” (Baudis et al., 2016) => RNN-CNN combo: 67.2

Goal-oriented Dialog (Bordes et al., arxiv16)



- bAbI tasks for goal-oriented dialog
- 6 tasks around restaurant booking
- Involve manipulate language and KB symbols

MemNNs for Goal-oriented Dialog

Time	Locutor	Dialog History	Hop #1	Hop #2	Hop #3
1	User	hi	.054	.016	.020
2	Bot	hello what can i help you with today	.040	.024	.008
3	User	may i have a table in paris	.099	.137	.141
4	Bot	i'm on it	.048	.028	.004
5	User	<silence>	.181	.099	.037
6	Bot	any preference on a type of cuisine	.056	.090	.014
7	User	i love indian food	.159	.188	.238
8	Bot	how many people would be in your party	.051	.022	.010
9	User	we will be six	.129	.144	.114
10	Bot	which price range are looking for	.039	.028	.006
11	User	in a moderate price range please	.128	.213	.405
12	Bot	ok let me look into some options for you	.016	.011	.003
User input		<silence>			
Correct answer		`api_call indian paris six moderate`			
Predicted answer		`api_call indian paris six moderate` **[Correct]**			

Next Steps

Variants of the class…

Some options and extensions:

- **Representation of inputs and memories could use all kinds of encodings:** bag of words, RNN style reading at word or character level, etc.
- **Different possibilities for output module:** e.g. multi-class classifier or uses an RNN to output sentences.
- **If the memory is huge** (e.g. Wikipedia) we need to organize the memories. Solution: hash the memories to store in buckets (topics). Then, memory addressing and reading doesn't operate on *all* memories.
- **If the memory is full**, there could be a way of removing one it thinks is most useless; *i.e.* it ``forgets'' somehow. That would require a scoring function of the utility of each memory..

Conclusion

1. (Key-Value) Memory Networks: promising model for jointly using symbolic and continuous systems
   - Can be trained end-to-end through backpropagation + SGD
   - Provide a great flexibility on how to design memories

2. bAbI, CBT, MovieQA, etc.: new tools for developing learning algorithms
   - Training and evaluation sets of reasonable sizes
   - Designed to ease interpretation

Open Research

- Papers:
  - Key-Value Memory Networks: http://arxiv.org/abs/1606.03126
  - Memory Networks: http://arxiv.org/abs/1410.3916
  - End-to-end Memory Networks: http://arxiv.org/abs/1503.08895
  - bAbI tasks: http://arxiv.org/abs/1502.05698
  - Children's Books Test: http://arxiv.org/abs/1511.023701
  - Large-scale QA with Memory Networks: http://arxiv.org/abs/1506.02075
  - Evaluating pre-requisite qualities of dialog systems: http://arxiv.org/abs/1511.06931
  - Dialog bAbI tasks: http://arxiv.org/abs/1605.07683
  - Dialog-based language learning: http://arxiv.org/abs/1604.06045

MemNN

Q&A

Dialog

- Data: fb.ai/babi (7 datasets including bAbI tasks, CBT and MovieQA)
- Code:
  - Memory Networks: https://github.com/facebook/MemNN
  - bAbI tasks generator: https://github.com/facebook/bAbI-tasks

RepEval @ ACL 2016

RepEval 2016

The First Workshop on Evaluating Vector Space Representations for NLP

12th August 2016, Berlin, Germany

Mission Statement: *To develop new and improved ways of measuring the quality or understanding the properties of vector-space representations in NLP.*

https://sites.google.com/site/repevalacl16/