Memory Networks for Language Understanding

Antoine Bordes - Facebook AI Research
LXMLS – Lisbon July 28, 2016
Bots?

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
2. Creation (and release) of datasets for training/evaluating those
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 = $\text{max} \Rightarrow \text{Hard attention}$
- requires explicit supervision of attention during training
- Only feasible for simple tasks

![Diagram of Memory Networks]

- Memory vectors (unordered)
- Controller module
- Internal state vector
- Supervision
- Read
- Addressing

\[
\{\vec{m}_1, \vec{m}_2, \ldots, \vec{m}_X, \ldots\}
\]
End-to-end Memory Networks (Sukhbaatar et al., NIPS15)

Memory Module

Controller module

Input

Output supervision
Memory Module

To controller (added to controller state)

Addressing signal (controller state vector)

Memory vectors

Attention weights / Soft address
Memory Networks on bAbI

Knowledge Source

Question

Where is the milk?

John dropped the milk. John took the milk there. Sandra went to the bathroom. John moved to the hallway. Mary went back to the bedroom.

m1: John dropped the milk.
m2: Sandra went to the bathroom.
m3: John took the milk there.
m4: John moved to the hallway.
m5: Mary went back to the bedroom.

Where is the milk?
## Dashboard

<table>
<thead>
<tr>
<th>TASK</th>
<th>N-grams</th>
<th>LSTMs</th>
<th>MemN2N</th>
<th>Memory Networks</th>
<th>StructSVM+coref+srl</th>
<th>Training on 1k stories</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1. Single supporting fact</td>
<td>36</td>
<td>50</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
</tr>
<tr>
<td>T2. Two supporting facts</td>
<td>2</td>
<td>20</td>
<td>87</td>
<td>PASS</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>T3. Three supporting facts</td>
<td>7</td>
<td>20</td>
<td>60</td>
<td>PASS</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>T4. Two arguments relations</td>
<td>50</td>
<td>61</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T5. Three arguments relations</td>
<td>20</td>
<td>70</td>
<td>87</td>
<td>PASS</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>T6. Yes/no questions</td>
<td>49</td>
<td>48</td>
<td>92</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T7. Counting</td>
<td>52</td>
<td>49</td>
<td>83</td>
<td>85</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>T8. Sets</td>
<td>40</td>
<td>45</td>
<td>90</td>
<td>91</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>T9. Simple negation</td>
<td>62</td>
<td>64</td>
<td>87</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T10. Indefinite knowledge</td>
<td>45</td>
<td>44</td>
<td>85</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T11. Basic coreference</td>
<td>29</td>
<td>72</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T12. Conjunction</td>
<td>9</td>
<td>74</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T13. Compound coreference</td>
<td>26</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T14. Time reasoning</td>
<td>19</td>
<td>27</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T15. Basic deduction</td>
<td>20</td>
<td>21</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>T16. Basic induction</td>
<td>43</td>
<td>23</td>
<td>PASS</td>
<td>PASS</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>T17. Positional reasoning</td>
<td>46</td>
<td>51</td>
<td>49</td>
<td>65</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>T18. Size reasoning</td>
<td>52</td>
<td>52</td>
<td>89</td>
<td>PASS</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>T19. Path finding</td>
<td>0</td>
<td>8</td>
<td>7</td>
<td>36</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>T20. Agent's motivation</td>
<td>76</td>
<td>91</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td></td>
</tr>
</tbody>
</table>
# Attention during memory lookups

## Samples from toy QA tasks

### Story 1: 1 supporting fact

<table>
<thead>
<tr>
<th>Action</th>
<th>Support</th>
<th>Hop 1</th>
<th>Hop 2</th>
<th>Hop 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniel went to the bathroom.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Mary travelled to the hallway.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>John went to the bedroom.</td>
<td>0.37</td>
<td>0.02</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>John travelled to the bathroom.</td>
<td>0.60</td>
<td>0.98</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Mary went to the office.</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

**Where is John?** Answer: bathroom  Prediction: bathroom

### Story 2: 2 supporting facts

<table>
<thead>
<tr>
<th>Action</th>
<th>Support</th>
<th>Hop 1</th>
<th>Hop 2</th>
<th>Hop 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>John dropped the milk.</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>John took the milk there.</td>
<td>yes</td>
<td>0.88</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Sandra went back to the bathroom.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>John moved to the hallway.</td>
<td>yes</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Mary went back to the bedroom.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

**Where is the milk?** Answer: hallway  Prediction: hallway

### Story 3: 16 basic induction

<table>
<thead>
<tr>
<th>Character</th>
<th>Fact</th>
<th>Support</th>
<th>Hop 1</th>
<th>Hop 2</th>
<th>Hop 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brian</td>
<td>is a frog.</td>
<td>yes</td>
<td>0.00</td>
<td>0.98</td>
<td>0.00</td>
</tr>
<tr>
<td>Lily</td>
<td>is gray.</td>
<td>yes</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Brian</td>
<td>is yellow.</td>
<td>yes</td>
<td>0.07</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Julius</td>
<td>is green.</td>
<td>yes</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Greg</td>
<td>is a frog.</td>
<td>yes</td>
<td>0.76</td>
<td>0.02</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**What color is Greg?** Answer: yellow  Prediction: yellow

### Story 4: size reasoning

<table>
<thead>
<tr>
<th>Fact</th>
<th>Support</th>
<th>Hop 1</th>
<th>Hop 2</th>
<th>Hop 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>The suitcase is bigger than the chest.</td>
<td>yes</td>
<td>0.00</td>
<td>0.88</td>
<td>0.00</td>
</tr>
<tr>
<td>The box is bigger than the chocolate.</td>
<td>yes</td>
<td>0.04</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>The chest is bigger than the chocolate.</td>
<td>yes</td>
<td>0.17</td>
<td>0.07</td>
<td>0.90</td>
</tr>
<tr>
<td>The chest fits inside the container.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>The chest fits inside the box.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

**Does the suitcase fit in the chocolate?** Answer: no  Prediction: no

## Test Accuracy and Failed Tasks

<table>
<thead>
<tr>
<th>Model</th>
<th>Test Acc</th>
<th>Failed tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MemNN</td>
<td>93.3%</td>
<td>4</td>
</tr>
<tr>
<td>LSTM</td>
<td>49%</td>
<td>20</td>
</tr>
<tr>
<td>MemN2N</td>
<td>74.82%</td>
<td>17</td>
</tr>
<tr>
<td>1 hop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 hops</td>
<td>84.4%</td>
<td>11</td>
</tr>
<tr>
<td>3 hops</td>
<td>87.6%</td>
<td>11</td>
</tr>
</tbody>
</table>

20 bAbI Tasks
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.
- “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)

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 greatcoat, and four boys were prancing about him clambering 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 | archie | 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),'
   the Gryphon replied very solemnly .
12. Alice was thoroughly puzzled .
13. `Does the (boots and shoes) !'
14. she repeated in a wondering tone .
15. `Why , what are YOUR shoes done with ?'
16. `said the Gryphon .'
17. `I mean , what makes them so shiny ?'
18. Alice looked down at them , and considered a little before she gave
   her answer .
19. 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

<table>
<thead>
<tr>
<th>Methods</th>
<th>Named Entities</th>
<th>Common Nouns</th>
<th>Verbs</th>
<th>Prepositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humans (query) (*)</td>
<td>0.320</td>
<td>0.644</td>
<td>0.716</td>
<td>0.676</td>
</tr>
<tr>
<td>Humans (context+query) (*)</td>
<td><strong>0.816</strong></td>
<td><strong>0.816</strong></td>
<td><strong>0.828</strong></td>
<td>0.708</td>
</tr>
<tr>
<td>Maximum frequency (corpus)</td>
<td>0.120</td>
<td>0.158</td>
<td>0.373</td>
<td>0.315</td>
</tr>
<tr>
<td>Maximum frequency (context)</td>
<td>0.335</td>
<td>0.281</td>
<td>0.285</td>
<td>0.275</td>
</tr>
<tr>
<td>Sliding window</td>
<td>0.168</td>
<td>0.196</td>
<td>0.182</td>
<td>0.101</td>
</tr>
<tr>
<td>Word distance model</td>
<td>0.398</td>
<td>0.364</td>
<td>0.380</td>
<td>0.237</td>
</tr>
<tr>
<td>Kneser-Ney language model</td>
<td>0.390</td>
<td>0.544</td>
<td>0.778</td>
<td>0.768</td>
</tr>
<tr>
<td>Kneser-Ney language model + cache</td>
<td>0.439</td>
<td>0.577</td>
<td>0.772</td>
<td>0.679</td>
</tr>
<tr>
<td>Embedding model (context+query)</td>
<td>0.253</td>
<td>0.259</td>
<td>0.421</td>
<td>0.315</td>
</tr>
<tr>
<td>Embedding model (query)</td>
<td>0.351</td>
<td>0.400</td>
<td>0.614</td>
<td>0.535</td>
</tr>
<tr>
<td>Embedding model (window)</td>
<td>0.362</td>
<td>0.415</td>
<td>0.637</td>
<td>0.589</td>
</tr>
<tr>
<td>Embedding model (window+position)</td>
<td>0.402</td>
<td>0.506</td>
<td>0.736</td>
<td>0.670</td>
</tr>
<tr>
<td>LSTMs (query)</td>
<td><strong>0.408</strong></td>
<td><strong>0.541</strong></td>
<td><strong>0.813</strong></td>
<td><strong>0.802</strong></td>
</tr>
<tr>
<td>LSTMs (context+query)</td>
<td><strong>0.418</strong></td>
<td><strong>0.560</strong></td>
<td><strong>0.818</strong></td>
<td><strong>0.791</strong></td>
</tr>
<tr>
<td>Contextual LSTMs (window context)</td>
<td>0.436</td>
<td>0.582</td>
<td>0.805</td>
<td><strong>0.806</strong></td>
</tr>
<tr>
<td>MemNNs (lexical memory)</td>
<td>0.431</td>
<td>0.562</td>
<td>0.798</td>
<td>0.764</td>
</tr>
<tr>
<td>MemNNs (window memory)</td>
<td>0.493</td>
<td>0.554</td>
<td>0.692</td>
<td>0.674</td>
</tr>
<tr>
<td>MemNNs (sentential memory + PE)</td>
<td>0.318</td>
<td>0.305</td>
<td>0.502</td>
<td>0.326</td>
</tr>
<tr>
<td>MemNNs (window memory + self-sup.)</td>
<td><strong>0.666</strong></td>
<td><strong>0.630</strong></td>
<td>0.690</td>
<td>0.703</td>
</tr>
</tbody>
</table>
We evaluate our models on the data from: “Teaching Machines to Read and Comprehend” Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will

<table>
<thead>
<tr>
<th>METHODS</th>
<th>VALIDATION</th>
<th>TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAXIMUM FREQUENCY (ARTICLE) (*)</td>
<td>0.305</td>
<td>0.332</td>
</tr>
<tr>
<td>SLIDING WINDOW</td>
<td>0.005</td>
<td>0.006</td>
</tr>
<tr>
<td>WORD DISTANCE MODEL (*)</td>
<td>0.505</td>
<td>0.509</td>
</tr>
<tr>
<td>DEEP LSTMs (ARTICLE+QUERY) (*)</td>
<td>0.550</td>
<td>0.570</td>
</tr>
<tr>
<td>CONTEXTUAL LSTMs (“ATTENTIVE READER”) (*)</td>
<td>0.616</td>
<td>0.630</td>
</tr>
<tr>
<td>CONTEXTUAL LSTMs (“IMPATIENT READER”) (*)</td>
<td>0.618</td>
<td>0.638</td>
</tr>
<tr>
<td>MemNNs (WINDOW MEMORY)</td>
<td>0.580</td>
<td>0.606</td>
</tr>
<tr>
<td>MemNNs (WINDOW MEMORY + SELF-SUP.)</td>
<td>0.634</td>
<td>0.668</td>
</tr>
<tr>
<td>MemNNs (WINDOW MEMORY + ENSEMBLE)</td>
<td>0.612</td>
<td>0.638</td>
</tr>
<tr>
<td>MemNNs (WINDOW MEMORY + SELF-SUP. + ENSEMBLE)</td>
<td>0.649</td>
<td>0.684</td>
</tr>
<tr>
<td>MemNNs (WINDOW + SELF-SUP. + ENSEMBLE + EXCLUD. COOCURRENCES)</td>
<td><strong>0.662</strong></td>
<td><strong>0.694</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Answer</th>
<th>ent193</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oisin Tymon</td>
<td></td>
</tr>
</tbody>
</table>
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

  
  QACNN: 77.4  CBT-NE: 71.9  CBT-CN: 69.0
Open Question Answering
Open-domain Question Answering

Answer questions on any topic

Knowledge Base (KB)

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 problem!

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
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.

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)

What year was the movie Blade Runner released?

KB
[Blade Runner, directed_by, Ridley Scott]
[Blade Runner, written_by, Philip K. Dick, Hampton Fancher] [Steven Spielberg, directed, Jurassic Park, …]
[Blade Runner, release_year, 1982]
[Blade Runner, has_tags, dystopian, noir, police, androids, …]

[139x903]Memory Networks for QA from KB (Bordes et al., arxiv15)

Memories ($m_1, m_2, m_3, m_4, …$)

[Blade Runner, directed_by, Ridley Scott]
[Blade Runner, written_by, Philip K. Dick, Hampton Fancher] [Steven Spielberg, directed, Jurassic Park, …]
[Blade Runner, release_year, 1982]
[Blade Runner, has_tags, dystopian, noir, police, androids, …]

What year was the movie Blade Runner released?

KB
[Blade Runner, directed_by, Ridley Scott]
[Blade Runner, written_by, Philip K. Dick, Hampton Fancher] [Steven Spielberg, directed, Jurassic Park, …]
[Blade Runner, release_year, 1982]
[Blade Runner, has_tags, dystopian, noir, police, androids, …]

[139x903]Memory Networks for QA from KB (Bordes et al., arxiv15)

Memories ($m_1, m_2, m_3, m_4, …$)

[Blade Runner, directed_by, Ridley Scott]
[Blade Runner, written_by, Philip K. Dick, Hampton Fancher] [Steven Spielberg, directed, Jurassic Park, …]
[Blade Runner, release_year, 1982]
[Blade Runner, has_tags, dystopian, noir, police, androids, …]

What year was the movie Blade Runner released?

KB
[Blade Runner, directed_by, Ridley Scott]
[Blade Runner, written_by, Philip K. Dick, Hampton Fancher] [Steven Spielberg, directed, Jurassic Park, …]
[Blade Runner, release_year, 1982]
[Blade Runner, has_tags, dystopian, noir, police, androids, …]
Memory Networks for QA from Text (Hill et al., ICLR16)

What year was the movie Blade Runner released?

Wikipedia Entry: Blade Runner
Blade Runner is a 1982 American neo-noir dystopian science fiction film directed by Ridley Scott and written by Hampton Fancher and David Peoples. 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…

1982

Tron
1982
police
Tom Cruise
…
Memory Networks on MovieQA

Standard QA System on KB: 93.5%

No Knowledge (embeddings):
- KB: 50.2%
- IE: 60.2%
- Wikipedia: 52.5%

Improvement: 24%
Structuring Memories

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

[Blade Runner, directed_by, Ridley Scott]
[Blade Runner, release_year, 1982]

directed by Ridley Scott and 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?
What year was the movie Blade Runner released?

1982
What year was the movie Blade Runner released?

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…
Results on MovieQA

<table>
<thead>
<tr>
<th>System on KB</th>
<th>93.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Knowledge (embeddings)</td>
<td>54.4%</td>
</tr>
</tbody>
</table>

- Standard QA System on KB: 93.5%
- No Knowledge (embeddings): 54.4%

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

Key-Value Memory Networks
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 small
- Word embeddings pre-trained
- Dropout regularization

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

<table>
<thead>
<tr>
<th>Methods</th>
<th>Validation (HITS@1)</th>
<th>Test (HITS@1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR†</td>
<td>N/A</td>
<td>48.81</td>
</tr>
<tr>
<td>RNN†</td>
<td>N/A</td>
<td>37.91</td>
</tr>
<tr>
<td>LSTM†</td>
<td>N/A</td>
<td>55.22</td>
</tr>
<tr>
<td>MemN2N 1-HOP</td>
<td>57.23</td>
<td>56.25</td>
</tr>
<tr>
<td>MemN2N 2-HOPS</td>
<td>64.28</td>
<td>63.51</td>
</tr>
<tr>
<td>MemN2N 3-HOPS</td>
<td>64.31</td>
<td>63.72</td>
</tr>
<tr>
<td>MemN2N 4-HOPS</td>
<td>64.01</td>
<td>62.82</td>
</tr>
</tbody>
</table>

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

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

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

• 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/