MEMORY BASED MODEL

A OVERVIEW

LUJUN ZHAO

2016/03/10

School of Mathematical Sciences
Fudan University
A **memory network** consists of a memory $m$ and four components $I, G, O$ and $R$ as follows:

- $I$: (input feature map) – converts the incoming input to the internal feature representation.
- $G$: (generalization) – updates old memories given the new input.
- $O$: (output feature map) – produces a new output, given the memories.
- $R$: (response) – converts the output into the response format desired.
Experiments

- Large-scale QA
  Dataset: ReVerb ClueWeb09 Extractions (14M statements, stored as (subject, relation, object) triples)

<table>
<thead>
<tr>
<th>Method</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Fader et al., 2013)</td>
<td>0.54</td>
</tr>
<tr>
<td>(Bordes et al., 2014b)</td>
<td>0.73</td>
</tr>
<tr>
<td>MemNN (embedding only)</td>
<td>0.72</td>
</tr>
<tr>
<td>MemNN (with BoW features)</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Figure: Results on the large-scale QA task
Experiments

- Simulated world QA
  Dataset: generated 7k statements and 3k questions from the simulator

- Combining simulated data and large-scale QA
Neural Turing Machines [Graves et al. 2014]

Figure: Neural Turing Machine Architecture
Experiments

- Copy
  Dataset: unknown (sequence of random binary vectors)

- Repeat Copy

- Associative Recall

- Dynamic N-Grams

- Priority Sorts
End-To-End Memory Networks [Sukhbaatar et al. 2015]

Figure: End-To-End Memory Networks
Experiments

- Synthetic Question and Answering
  Dataset: Facebook bAbI Dataset

- Language Modeling
  Dataset: Penn Tree Bank and Text8
Dynamic Memory Networks [Kumar et al. 2015]

Figure: Dynamic Memory Networks
Experiments

- **Question Answering**
  Dataset: *Facebook bAbI Dataset*

- **Sentiment Analysis**
  Dataset: *Stanford Sentiment Treebank*

- **Part-of-Speech Tagging**
  Dataset: *Wall Street Journal dataset*
Figure: Neural network extended with push-down stack
Experiments

- Learning simple algorithmic patterns
  Dataset: generated by simple algorithms

- Language modeling
  Dataset: Penn Tree Bank

<table>
<thead>
<tr>
<th>Model</th>
<th>Ngram</th>
<th>Ngram + Cache</th>
<th>RNN</th>
<th>LSTM</th>
<th>SRCN [24]</th>
<th>Stack RNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation perplexity</td>
<td>-</td>
<td>-</td>
<td>137</td>
<td>120</td>
<td>120</td>
<td>124</td>
</tr>
<tr>
<td>Test perplexity</td>
<td>141</td>
<td>125</td>
<td>129</td>
<td>115</td>
<td>115</td>
<td>118</td>
</tr>
</tbody>
</table>

Figure: Comparison of RNN, LSTM, SRCN and Stack RNN on language modeling task
Figure: Illustrating a Neural Stack’s Operations, Recurrent Structure, and Control
Experiments

○ **Synthetic Transduction Tasks**
  Dataset: randomly generated from a vocabulary of 128 meaningless symbols
  ○ Sequence Copying
  ○ Sequence Reversal
  ○ Bigram flipping

○ **ITG Transduction Tasks**
  Dataset: the source and target sequence are jointly generated by Inversion Transduction Grammars
Structured-Memory NTMs [Zhang et al. 2015]

Figure: NTM and NTM variants that use LSTM as controllers
The experiments is to show the convergence speed and quality of those three variants, compared to the NTM setting.

- **Copy Task**
  Dataset: randomly generated

- **Associative Recall Task**
  Dataset: randomly generated
Figure: Evolvable Neural Turing Machine
Experiments

- Copy Task
  Dataset: unknown

- T-Maze Task
  Dataset: unknown


S. Sukhbaatar, J. Weston, R. Fergus, et al.
End-to-end memory networks.

J. Weston, S. Chopra, and A. Bordes.
Memory networks.

W. Zhang and Y. Yu.
Structured memory for neural turing machines.