Answering Questions on the bAbI Dataset Using Memory Networks (with LSTMs)

John Miller, Vincent Su, Jack Zhu

Introduction

- Task: Answer a question given a set of facts
  - e.g. John went to the bathroom. Mary went to the Kitchen. Where is John?
- Dataset: Facebook bAbI dataset
  - Set of 20 synthetic tasks which serve as baseline for any 'complete' AI

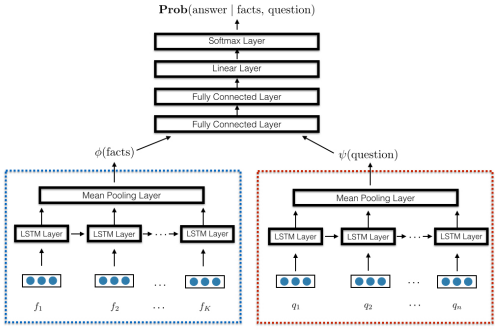
Model

- Memory Network
  - Idea: Combine an inference model with a database of knowledge for longer term storage
  - LSTM for inference model
    - Outputs a probability over words from the sentences and questions
  - LSTM for determining queries to knowledge database
    - Selects 'relevant' sentences

- Pre-trained word vectors replace words
  - Using the 50 dimensional GloVe representation
- Goal: Minimize combined objective of the sum of negative log likelihoods from the inference model and the memory access
  - i.e. the sum of NLL for selecting relevant sentences and predicting the write answer
  - In practice, Theano takes care of symbolic gradients

Data

- Each training example in the dataset is a "story"
- Stories are comprised of sentences and questions
- The training set has the correct answer to the question as well as indices for the sentences which are relevant to the answer



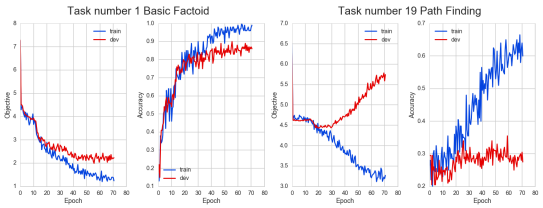
LSTM Layers figure out representations of the sentences and questions, which then feed into another neural network to output probabilities of answers

Results (example)

```
Supporting Facts:
sandra went to the kitchen .
john journeyed to the kitchen .
sandra went back to the garden .
mary travelled to the hallway .
sandra went back to the bathroom .
daniel journeyed to the garden .

Question:  where is daniel ?
True Answer:  garden
Predicted Answer:  garden
Predicted Relevant Facts:  john journeyed to the kitchen .
daniel journeyed to the garden .
```

Learning Curves



Conclusion/Future Work

- In summary, we implemented a memory network which outperforms a vanilla LSTM by selecting relevant facts to process
- Some of the bAbI tasks require higher order logic processing which was inaccessible to our model