Nicolas Drizard
nicolasdrizard@g.harvard.edu
CS 287

Due: March 25, 2016

Virgile Audi
vaudi@g.harvard.edu

Introduction

If we want to communicate and reason with a machine, then the machine will need to be able to ingest and understand the underlying logic of the sentences we communicate to it. Pick the Echo for instance, say you are to tell it that your mother just bought this great new phone on Amazon, and that it makes you jealous. Wouldn't it be great (or at least for Amazon) if the Echo understood that the answer to the question: "why am I jealous?" was the fact that you don't have the latest smartphone and replied by offering to order it for you immediately? This kind of tasks are called non-factoid question answering as they go beyond the scope of querying a knowledge base to answer a question such as "Who was the 1st President of the United States?". In this project, we would like to tackle the issue of non-factoid question answering by implementing the Dynamic Memory Network developped in [Kumar et al., 2015].

1 Problem Formulation

The goal of this project to build a model capable of solving the following type of problem:

1. Read a set of sentences forming a story

2. Given the story, the model answers a question by a word or a sentence itself

An example of such problems could be:

Story: John is bored. John goes out in the garden.
Q: Why is John in the garden?
A: John was bored

The model we plan on implementing is supposed to successfully complete a set of 20 task presented in [Weston et al., 2015].

Such tasks include:

- A single supporting fact fact, potentially amongst a set of other irrelevant facts, provides the answer;

- Two or three supporting facts that require to be chained to answer;

- Positional and size reasoning;

- Time reasoning, etc.

We will use Facebook's *bAbI* dataset which consists of 20 pairs of training and test sets of each 1000 questions.

2 Dynamic Memory Network

2.1 Presentation

To solve our problem, we will implement a recent variant of memory network [Weston et al., 2014] called Dynamic Memory Network (DMN). It is a neural network architecture which processes input sequences and questions, builds episodic memories and generates relevant answers. A question triggers an iterative attention process where the model will produce results while conditioning its attention on relevant information of the input to answer the question. The results are then reasoned over in a hierarchical recurrent sequence model to generate answers. This model can be used on several type of tasks: question answering, text classification for sentiment analysis and sequence modeling for part-of-speech tagging. We will focused mainly of the first task but if time will compare its performance on the other tasks with the methods implemented in the class.

2.2 Structure

The DMN is made up of four following modules that we present briefly. More documentation is provided in the original paper [Kumar et al., 2015]

- **Input Module**: Encodes raw text inputs (list of words or of sentences) into a vector representation;

- **Question Module**: Encodes the raw question (list of words) into a vector representation.;

- **Episodic Memory Module**: Uses an attention mechanism to choose which part of the inputs to focus on. It produces a "memory" vector representation taking into account the question as well as the previous memory; the memory is initialized to question. The iterative process gives the module the ability to retrieve new information, in the form of input representations, which were considered irrelevant in the previous iterations.

- **Answer Module**: Generates an answer from the final memory vector of the memory module.

2.3 Experimental Setup

Here we describe the implementation we plan to choose. First, each module uses a recurrent neural network to produce its output, built from the hidden states of the RNN. The paper mentions two alternatives to avoid suffering from the vanishing gradient problem: gated recurrent network [Cho et al., 2014] or long-short term memory [Hochreiter and Schmidhuber, 1997]. We will use the GRU as in the paper for computationnal reasons.Second, we will apply a transfer learning approach where we will initialize our embedding space to the pretrained GloVe [Pennington et al., 2014].Third, we will use the cross-entropy with an L2 regularization as the loss of our RNN.

3 Metric and Baseline

3.1 Metric

The metric used to evaluate the quality of the model will be the accuracy on the 20 test sets of the Facebook's *bAbI* dataset, as in the original paper. We will consider the project to be a success if we manage to reproduce the results presented in the paper, i.e. an average accuracy of 93.6% of the 20 *bAbi* tasks.

3.2 Baseline

A first baseline would be provided by a count based model using bag-of-words features from both the input and the question. We will try to use a simple LSTM which reads the paragraph text and a question, and outputs an answer word as another baseline also.

4 Possible Extension

One possible extension could be to apply our model on other dataset. The bAbI dataset has been built artificially and may lack of complexity. We could adapt our method to solve the QA tasks provided in the MCTest [Richardson et al.,].

Another possible extension is provided in the work of [Caiming Xiong, 2016] where the DMN architecture is improved when supporting facts are not marked during training.

References

[Caiming Xiong, 2016] Caiming Xiong, Stephen Merity, R. S. (2016). Dynamic memory networks for visual and textual question answering.

[Cho et al., 2014] Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259.

[Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory.

[Kumar et al., 2015] Kumar, A., Irsoy, O., Su, J., Bradbury, J., English, R., Pierce, B., Ondruska, P., Gulrajani, I., and Socher, R. (2015). Ask me anything: Dynamic memory networks for natural language processing. *CoRR*, abs/1506.07285.

[Pennington et al., 2014] Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

[Richardson et al.,] Richardson, M., Burges, C. J. C., and Renshaw, E. Mctest: A challenge dataset for the open-domain machine comprehension of text.

[Weston et al., 2015] Weston, J., Bordes, A., Chopra, S., and Mikolov, T. (2015). Towards ai-complete question answering: A set of prerequisite toy tasks. *CoRR*, abs/1502.05698.

[Weston et al., 2014] Weston, J., Chopra, S., and Bordes, A. (2014). Memory networks. *CoRR*, abs/1410.3916.