

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 is the type of task that goes outside the scope of Factoid Question Answering that queries a knowledge base to answer 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 Answer by implementing a Dynamic Memory Network for Natural Language Processing, introduced in "Ask Me Anything: Dynamic Memory Networks for Natural Language Processing" [1].

1 Problem Formulation

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

1. Read a set of sentences forming a story or knowledge base
2. Given the story, the model answers a question on this story 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 tasks presented in "Towards AI-complete question answering: a set of prerequisite toy tasks" by J. Weston et al. [2] Such tasks include:

- Single Supporting Fact: "consists of questions where a previously given single supporting fact, potentially amongst a set of other irrelevant facts, provides the answer"
- Two or Three Supporting Facts: "two supporting statements have to be chained to answer the question"
- Positional and Size Reasoning
- Time Reasoning, etc.

We will use the standard *bAbI* dataset which consists of 20 pairs of training and test sets of each 1000 questions.

2 Dynamic Memory Network

Present and explain the different modules (using GRU or LSTM)

3 Metric and Baseline

3.1 Metric

The metric we will use to evaluate the quality of the model will be a measure of the accuracy on the 20 test sets of the *bAbi* dataset. This is the same metric used in [1]. 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. If it is relatively easy to obtain positive results on the *bAbi* dataset, as it is claimed in [1] that the model can be applied to many other such as Part-of-speech tagging, we plan to benchmarked the performance of the model using the models developed in HW2.

3.2 Baseline

Count based model

4 Possible Extension

paper Visual Question Answering

5 Bibliography

1. *Ask Me Anything: Dynamic Memory Networks for Natural Language Processing*, A. Kumar, P. Ondruska, M. Iyyer, J. Bradbury, I. Gulrajani, V. Zhong, R. Paulus & R. Socher, MetaMind, 2016
2. *Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks*, J. Weston, A. Bordes, S. Chopra, A. M. Rush, B. van Merriënboer, A. Joulin, T. Mikolov, Facebook AI Research, 2015