Answering Reading Comprehension Using MemNNs

Darshan Kapashi, Pararth Shah

Introduction

- In this work, we explore the use of memory neural networks for answering multiple choice questions in reading comprehension datasets.
- MemNN (Weston et al [1]) is a recent model that aims to learn how to reason with inference components and a long-term memory component.
- We study the performance of memory networks on these datasets:
  - bAbI: Toy QA tasks introduced in Weston et al [2]
  - MCTest: Real-world reading comprehension dataset with multiple choice question answers [3]
  - Wiki-QA: Manually-generated factoid questions from a corpus of wikipedia articles [4]
- Models:
  - LSTM
  - Supervised MemNNs [1]
  - Weakly supervised MemNNs [5]

Sample bAbI QA Task

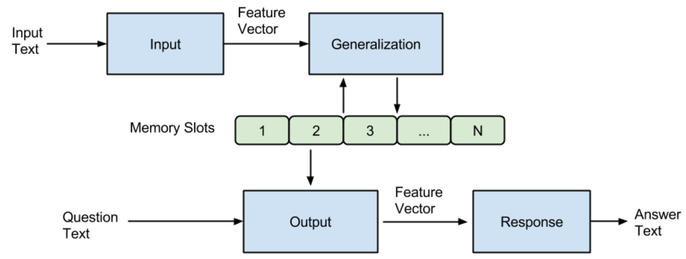
- **Input:** John is in the playground. Bob is in the office. John picked up the football. Bob went to the kitchen.
- Q: Where is the football? **A: playground**
- Q: Where was Bob before the kitchen? **A: office**

Sample MCTest Task

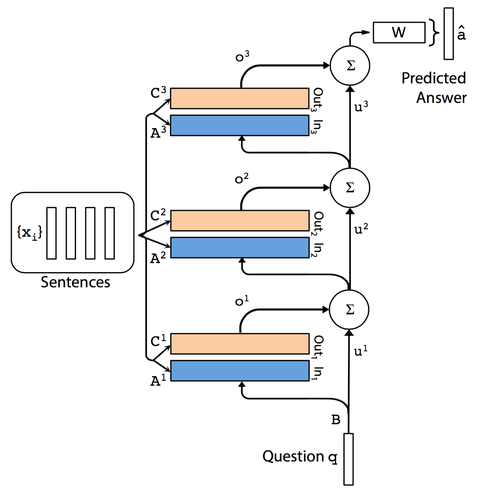
- **Input:** One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.
- Q: What did James pull off of the shelves in the grocery store?
- **A) pudding** B) fries C) food D) splinters
- Q: Where did James go after he went to the grocery store
- A) his deck B) his freezer **C) a fast food restaurant** D) his room
- Q) What did James do after he ordered the fries?
- A) went to the grocery store **B) went home without paying** C) ate them D) made up his mind to be better

MemNN Architecture

Supervised MemNN



Weakly Supervised MemNN



Pruning Memories For Efficiency

- Increase in vocabulary size takes longer to train the WMemNN.
- Increase in number of sentences per article also slows down the training time. For instance, the wiki-QA dataset has 500 statements per article on average.
- We attempted to prune the size of memory considered for prediction by two approaches:
  - Hash memories by named entities; only consider memories with at least 1 entity common with question
  - Sort memories by distance of average word vector with that of the question and pick top-k for consideration

Results

Test Set Accuracy (%)				
Dataset	**LSTM w/ relevant stmts**	**LSTM w/ entire input**	**Supervised MemNN (k=2)**	**Weakly sup. MemNN**
QA1	100	31.2	100	66.7
QA2	100	35.6	77.5	60.3
QA3	100	27.1	22.9	50.7
MCTest160	-	51.6	-	45.2
MCTest500	-	40.1	-	36.5

Why MC Questions Need A Different Approach

- Multiple choice questions can be formulated as a probability estimation over answer choices, instead of a word sequence generator (eg. LSTM) in the output.
- This implies that deep learning based approaches can handle more complex reading comprehension tasks, as long as the candidate answers are appropriately integrated into the model.
- The weakly supervised MemNN model can only output single word answers. One option is to attach an LSTM to the output of WMemNN.

Error Analysis (MCTest500)

Error 1:
When I woke up this morning, my cat was not inside the house. She is allowed outside, and **we have a cat door in the kitchen** so she can come and go as she pleases. … So it was very weird to wake up and see that she was not home for her morning food! Guess where she was? **Trapped in the garage** overnight!

Q: Where does my cat have a door?
A) Den **B) Garage** **C) Kitchen** D) Box

Error 2:
Steve loved that Bob would give him fish more than anything. **Steve also liked bananas, oranges and apples, but fish was his favorite**.

Q) What was Steve's favorite food?
A) bananas **B) fish** C) oranges **D) apples**

Contact

Pararth Shah (pararth@stanford.edu)
Darshan Kapashi (darshank@stanford.edu)

References

[1] Weston, Jason, Sumit Chopra, and Antoine Bordes. "Memory networks." arXiv preprint arXiv:1410.3916 (2014).
[2] Weston, Jason, et al. "Towards AI-complete question answering: a set of prerequisite toy tasks." arXiv preprint arXiv:1502.05698 (2015).
[3] Richardson, Matthew, Christopher JC Burges, and Erin Renshaw. "MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text." EMNLP. 2013.
[4] Smith, Noah A., Michael Heilman, and Rebecca Hwa. "Question generation as a competitive undergraduate course project." Proceedings of the NSF Workshop on the Question Generation…
[5] Sukhbaatar, Sainbayar, et al. "Weakly Supervised Memory Networks." arXiv preprint arXiv:1503.08895 (2015).