Memory Networks Model for Wikipedia QA

ASHWIN

University of Colorado Boulder

ashwin.asokan@colorado.edu

May 1, 2017

Problem Statement

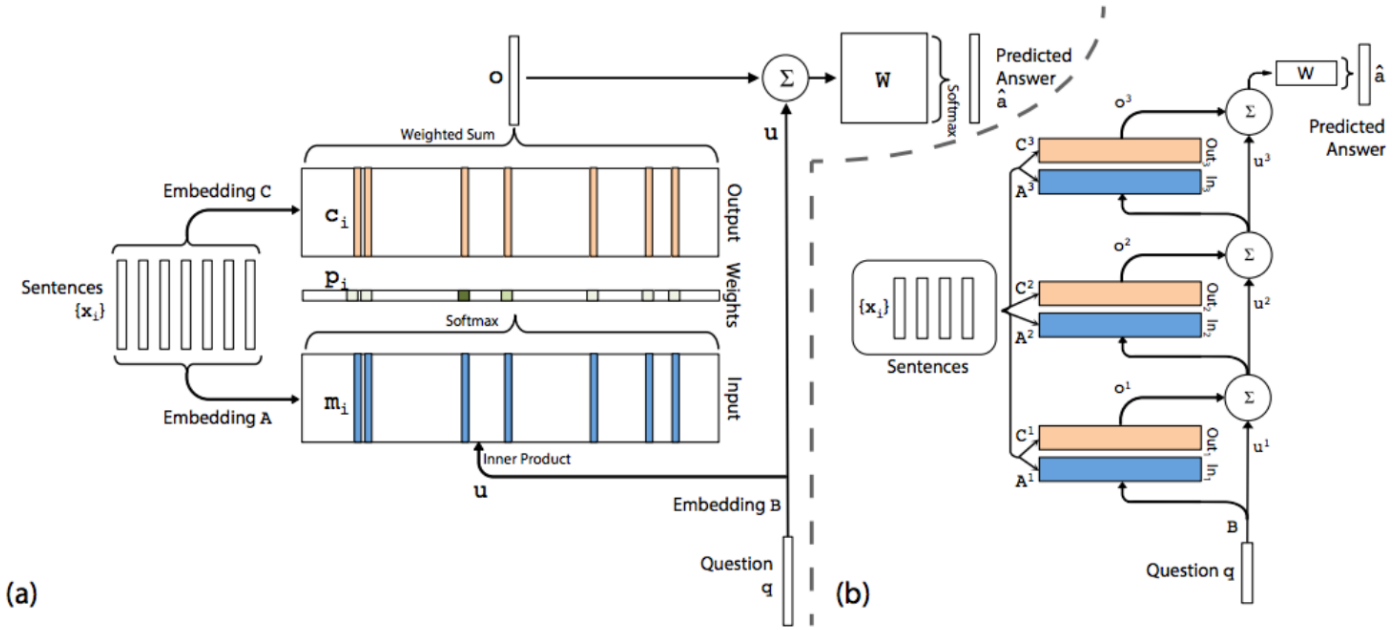
- Question answering given closed knowledge source is an useful and challenging NLP task
- Recently Memory neural network models has shown promising results for bAbI tasks,a synthetic dataset which requires inferring answers from hundreds of facts.
- My task is to apply that model to wikipedia dataset in which case no of facts to consider for a question is considerably more.
- The model should be able to select relevant part of the page based on the question and generate answers.
- Wikipedia data set has 16M traning samples, 1M validation set and another 1M test samples.

Example

Folkart Towers are twin skyscrapers in the Bayrakli district of the Turkish city of Izmir. Reaching a structural height of 200 m (656 ft) above ground level, they are the tallest . . .

Question country
Answer Turkey

End-To-End Memory Networks Sukhbaatar et al 2015



End-To-End Memory Networks Sukhbaatar et al 2015

Input Memory Representation

$$P(a|s,q)$$

$$sl = max(storylength); ql = max(querylength);$$

$$v = vocabsize; D = Embeddingdimension$$

$$m[sl, D] = Embeddings[x(sl), A(v, D)]$$

$$u[ql, D] = Embeddings[q(ql), B(v, D)]$$

$$p[sl, ql] = Softmax(u * m)$$

End-To-End Memory Networks Sukhbaatar et al 2015

Output Memory Representation

$$c[sl, ql] = Embeddings[x(sl), C(v, ql)]$$
$$o[sl, ql] = p * c$$

Generate Answers

$$a = Softmax(RNN([o, u]))$$

Parameters to learn

$$A, B, C, H$$

Notes

- Not able to train beyond 10K samples. Exhausts memory. 90M params to learn.
- Pruning the story by finding similarities between question and story line pairs and pick top k.
- To compute similarity between two sentence, I found something called word mover's distance. Its basically a word embedding distance measure between distribution of words between sentence pairs. - "From Word Embeddings To Document Distances" by Kusner et al 2015
- I have 15-20 percent validation accuracy in my runs, but Paper reports 80 percent for some categories.
- TODO: Parameter tunings, as much error analysis as possible.

References

- "End-To-End Memory Networks" by Sukhbaatar et al 2015
- "From Word Embeddings To Document Distances" by Kusner et al 2015
- "WIKI READING : A Novel Large-scale Language Understanding Task over Wikipedia" by Hewlett et al 2016