Memory Network For Question Answering

Virgile Audi
Nicolas Drizard

TF: Sam Wiseman

May, 3^{rd} 2015

Outline

Table of Contents

Supporting Facts

Story

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.

Q Where is Mary? **A** office

Story

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.

Q Where was the apple before the kitchen? **A** office

Reasoning

Story

Sheep are afraid of wolves.
Cats are afraid of dogs.
Mice are afraid of cats.
Gertrude is a sheep.

Q What is Gertrude afraid of? **A** wolves

Story

Lily is a swan.
Lily is white.
Bernhard is green.
Greg is a swan.

Q What color is Greg? **A** white

Intuition

How do Humans build their answer?

1. Type of question
2. Occurence of the words from the question
3. Associations of words (*memory*)
4. Meaning of words (*reasoning*, *interpretation*)

Table of Contents

Count Based Model

Prediction based on two features:

$$\hat{y}(X, Q) = argmax(\log(f_1(X)) + \log(f_2(Q)))$$

with

- (X, Q) tuple story, question
- $f_1(X)$: answer words counts in the story (weighted by order of appearance)
- $f_2(Q)$: embedding of the question based on possible answers question word

Table of Contents

Single Hop Architecture

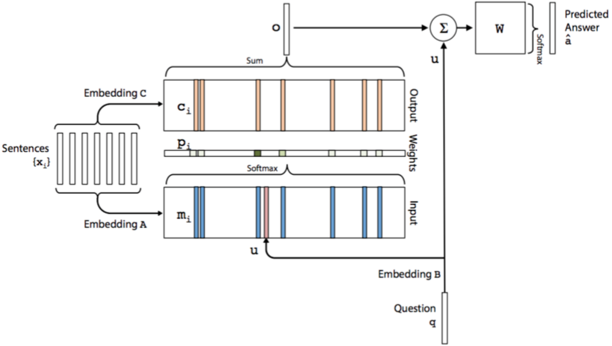

Figure: Single Hop architecture

Source: End-To-end Memory Networks

Multiple Hops Architecture

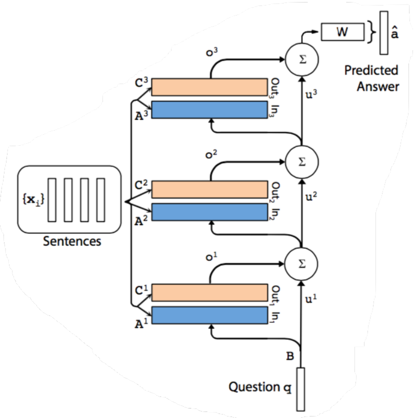

Figure: Multiple Hops architecture

Source: End-To-end Memory Networks

Parameters Tying

Two Approaches to reduce the number of parameters:

- **Adjacent**
  - $A^{k+1} = C^k$
  - $B = A^1$
- **RNN-like**
  - $A^1 = A^2 = ... = A^k$
  - $C^1 = C^2 = ... = C^k$
  - $u^{k+1} = Hu^k + o^k$

Implementation Tricks

- bag-of-words representation $x_i = \{x_{i1}, ..., x_{is}\}$ becomes $m_i = \sum_j Ax_{ij}$
- Temporal encoding $m_i = \sum_j Ax_{ij} + T_A(i)$
- high variance, best model over several training

Table of Contents

Table of Contents