Deep Learning for Nature Language Processing

Lianhao Yin

Lund University

lianhao.yin@energy.lth.se

November 29, 2016

Overview

Natural Language Processing

- Natural language processing is a field at the intersection of computer science artificial intelligence and linguistics.
- Goal: for computers to process or understand natural language in order to perform tasks that are useful: speech recognition, translation, answering questions and so on.

Reference

- http://cs224d.stanford.edu/syllabus.html
- https://www.youtube.com/watch?v=oGk1v1jQITw
- https://www.youtube.com/watch?v=ReUrmqStBd4
- https://deeplearning4j.org/word2vec

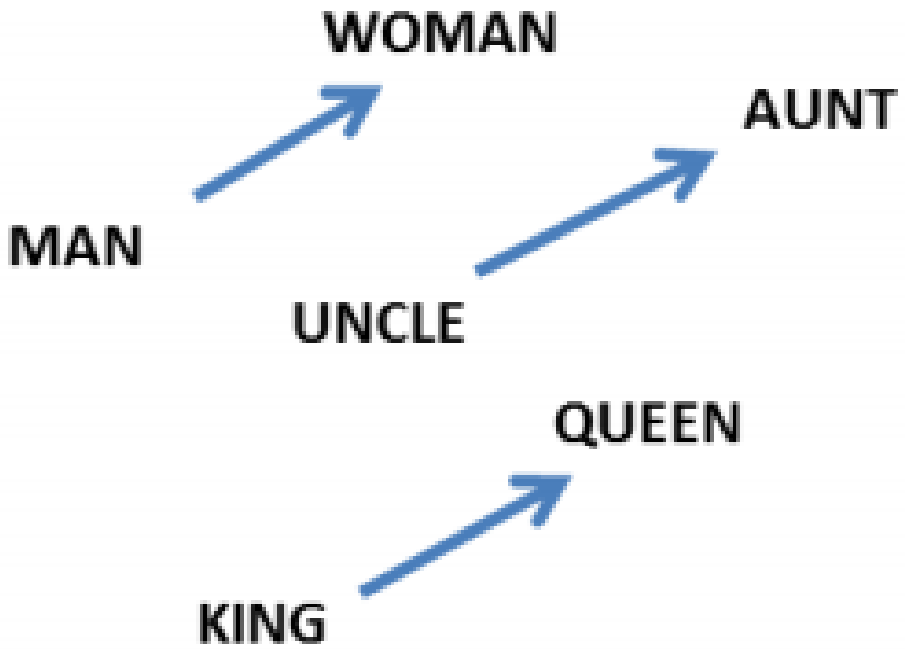
A basic problem in speech recognition

- We can not identify phonemes perfectly in noisy speech.
- People use their understanding of the meaning of the utterance to hear the right words
- This means the speech recognizers have to know which are likely to come next and which are not

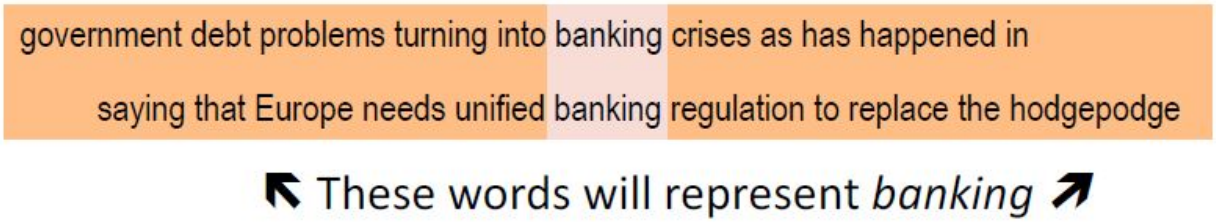
Word Embedding

wikipedia: Word embedding is the collective name for a set of language modeling and feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers

- $W(woman) - W(man) \cong W(aunt) - W(uncle)$
- $W(woman) - W(man) \cong W(queen) - W(king)$
- $W(aunt) - W(uncle) \cong W(queen) - W(king)$

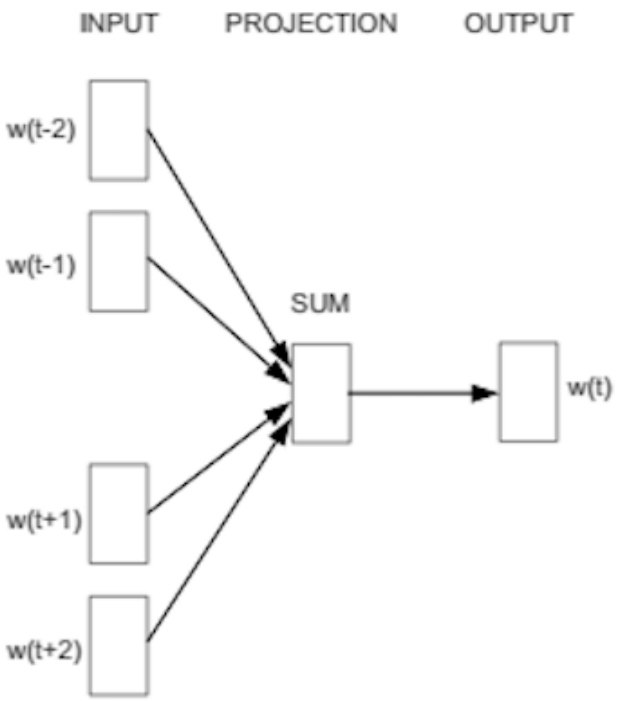


Representations for Language Tasks: Machine Translation

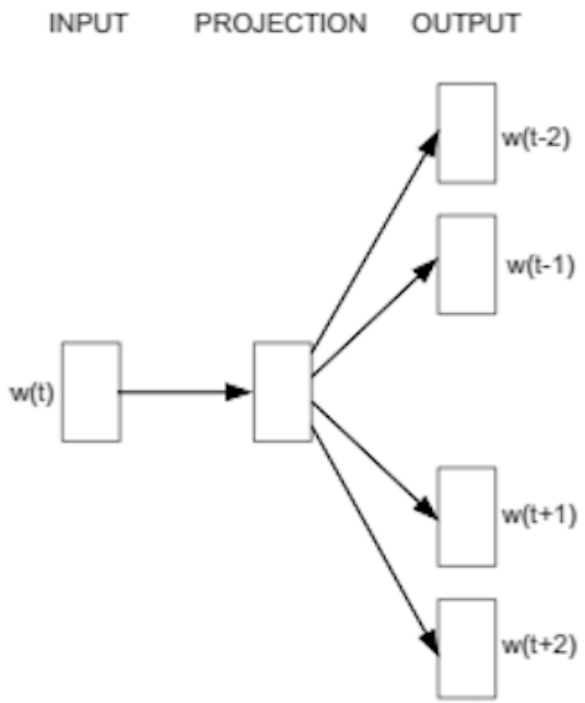


- You can get a lot of value by representing a word by means of its neighbors

Word to Vector



CBOW



Skip-gram

Details of Word2Vec

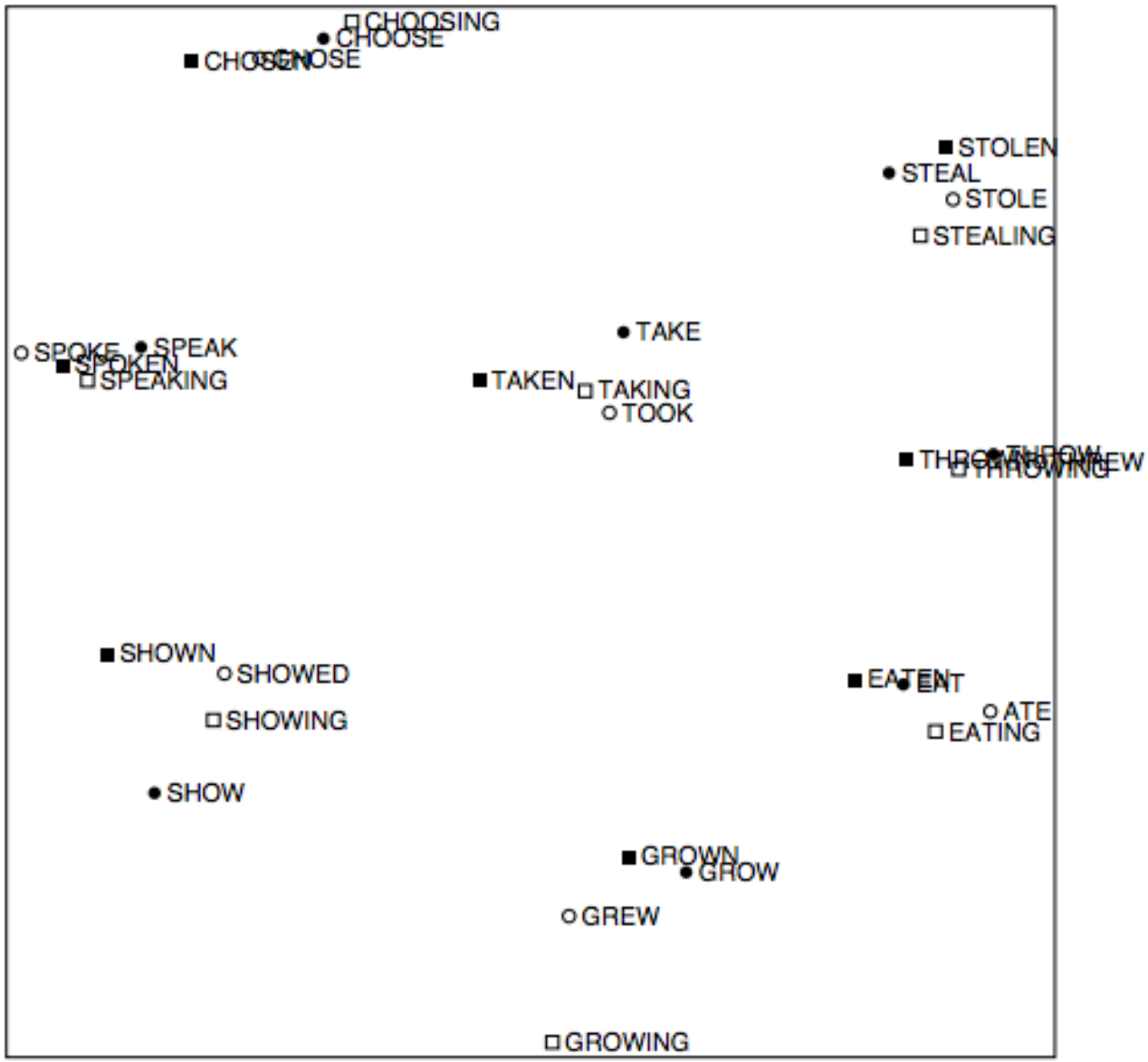
Objective function: Maximize the log probability of any context word given the current center word:

$$J(\theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\substack{-m \leqslant j \leqslant m \\ j \neq 0}} log\, p(w_{t+j}|w_t) \tag{1}$$

$$p(o|c) = \frac{exp(u_o^T v_c)}{\sum_{w=1}^{W} exp(u_w^T v_c)}$$

where o is the outside (or output) word id, c is the center word id, v and u are center and outside vectors of indices c and o

Similarity using word vector



GloVe (Global Vectors for Word Representation) results

GloVe is an unsupervised learning algorithm for obtaining vector representations for words

Nearest words to frog:

1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

litoria

leptodactylidae

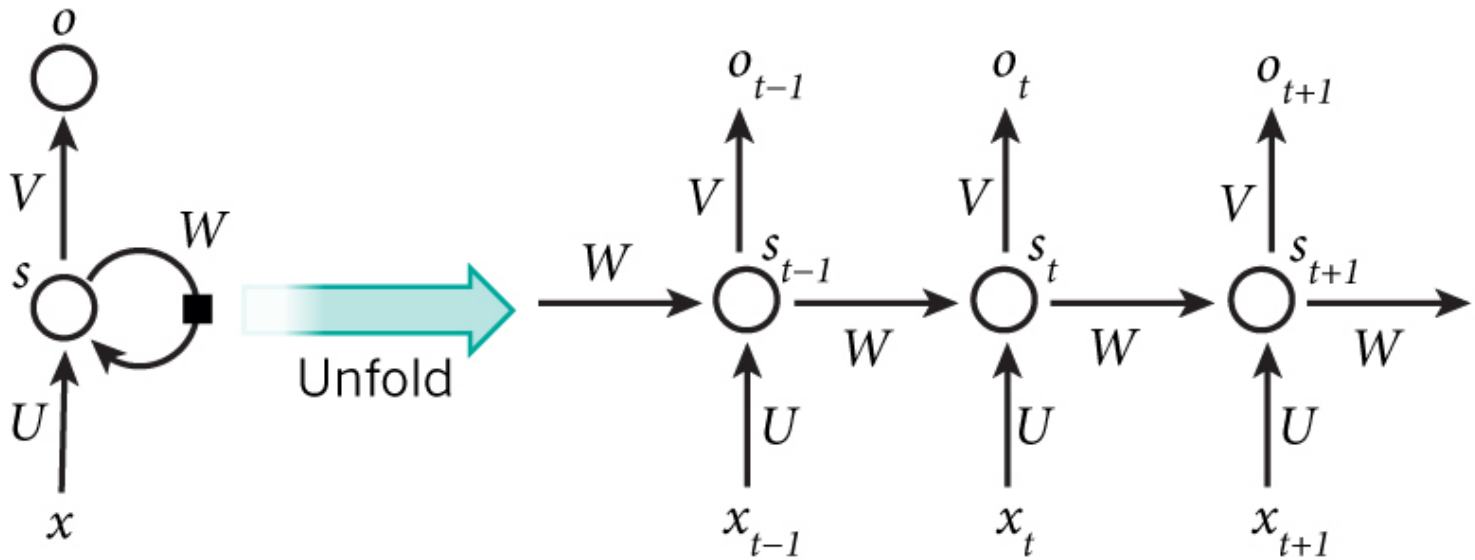
rana

eleutherodactylus

Ref:http://nlp.stanford.edu/projects/glove/

Recurrent Neural Network

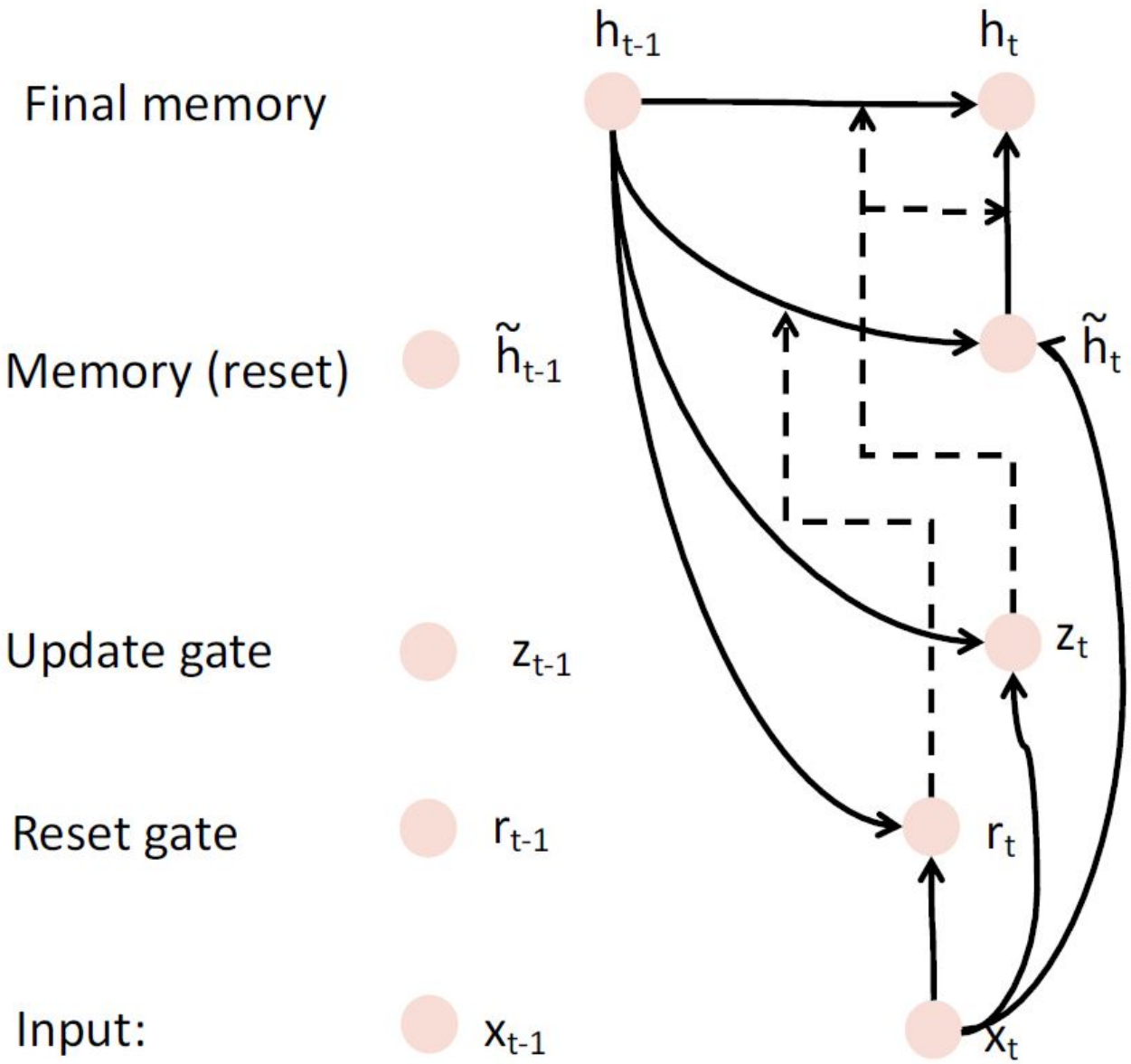
A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior.



Gated Recurrent Units (GRU)

- Gated Recurrent Units (GRU) introduced by Cho et al. 2014.
- Keep around memories to capture long distance dependencies
- Allow error messages to flow at different strengths depending on the inputs

Gated Recurrent Units (GRU)



Gated Recurrent Units (GRU)

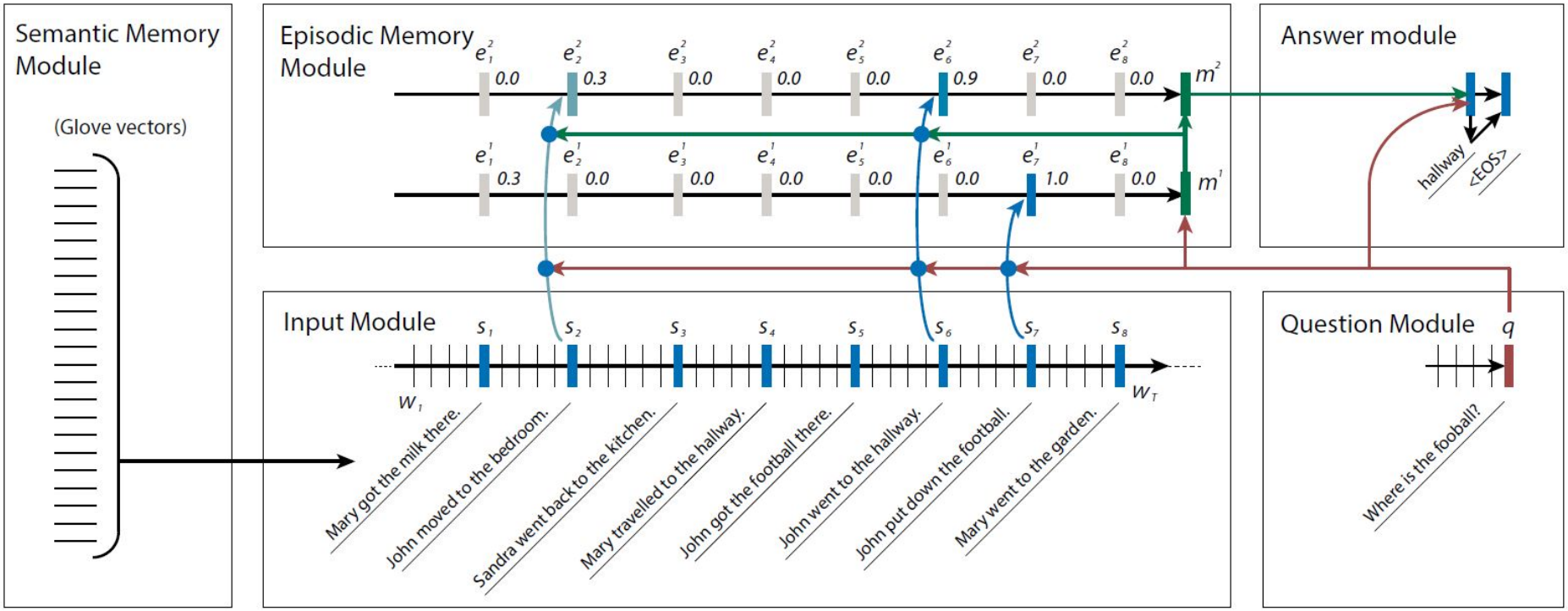
- Standard RNN computes hidden layer at next time step directly:
  $h_t = f(W_{h_{t-1}}^{(hh)} + W^{(hx)} x_t)$
- GRU first computes an update gate (another layer) based on current input word vector and hidden state
  $z_t = \sigma(W_{x_t}^{(z)} + U_{h_{t-1}}^{(z)})$
- Compute reset gate similarly but with different weights
  $r_t = \sigma(W_{x_t}^{(r)} + U_{h_{t-1}}^{(r)})$

High level idea for harder questions

- Imagine having to read an article, memorize it, then get asked various questions
- You can't store everything in working memory
- Optimal: give you the input data, give you the question, allow as many glances as possible

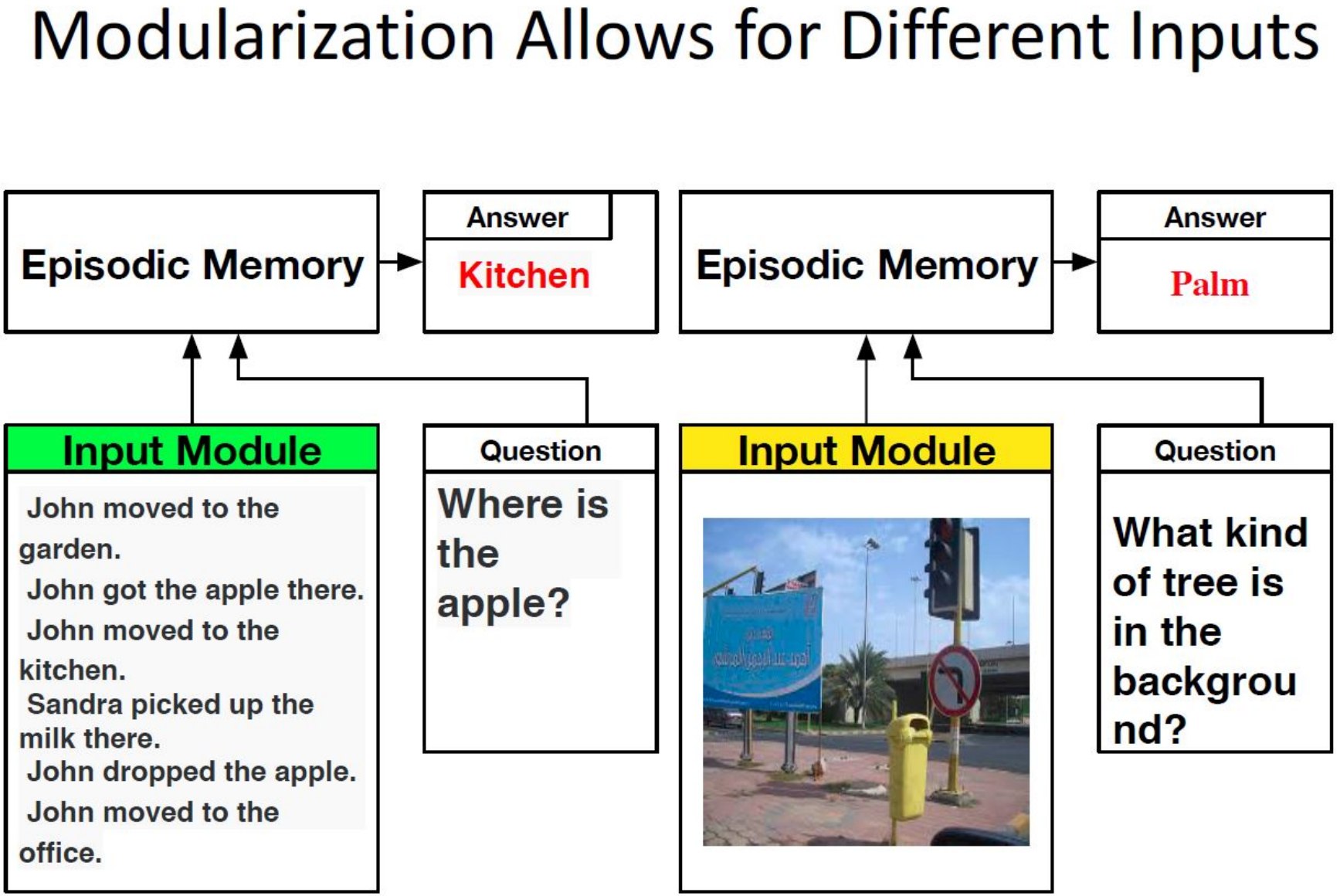
```
1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?        bathroom        1
4 Daniel went back to the hallway.
5 Sandra moved to the garden.
6 Where is Daniel?      hallway         4
7 John moved to the office.
8 Sandra journeyed to the bathroom.
9 Where is Daniel?      hallway         4
10 Mary moved to the hallway.
11 Daniel travelled to the office.
12 Where is Daniel?     office          11
13 John went back to the garden.
14 John moved to the bedroom.
15 Where is Sandra?     bathroom        8
1 Sandra travelled to the office.
2 Sandra went to the bathroom.
3 Where is Sandra?      bathroom        2
```

Dynamic Memory Networks



Ref: Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, Ankit Kumar. Source: https://arxiv.org/pdf/1506.07285v5.pdf

Modularization allows for different inputs



Attention visualization


What is the main color on the bus ?

Answer: blue


What type of trees are in the background ?

Answer: pine


How many pink flags are there ?

Answer: 2


Is this in the wild ?

Answer: no

Homework

try out the example of vector representations of words using tensor flow. https://www.tensorflow.org/versions/r0.11/tutorials/word2vec/index.html or improved-Dynamic-Memory-Networks-DMN-plus with theano https://github.com/ethancaballero/Improved-Dynamic-Memory-Networks-DMN-plus

The End