Deep Learning in NLP

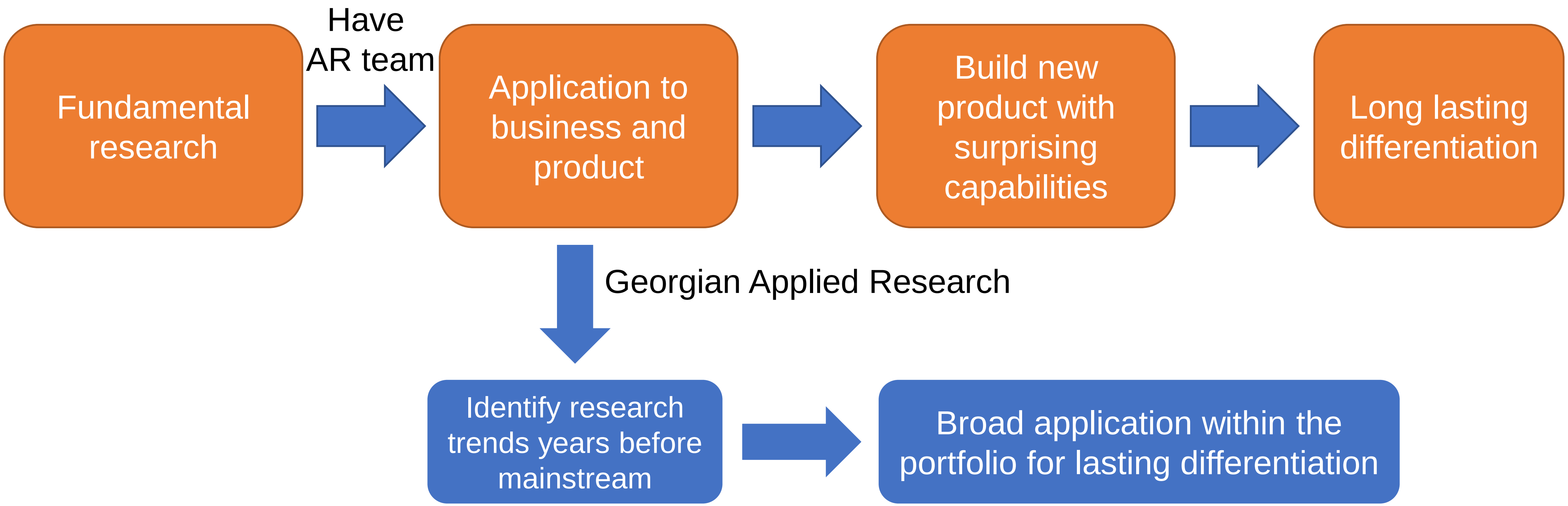
Parinaz Sobhani

Bio

- Director, Machine Learning at Georgian Partners
- Ph.D. Computer Science (University of Ottawa)
- Research: NLP, DL, Sentiment and Stance Classification
- Previous affiliations:
  - Research Intern: Microsoft Research
  - Research Intern: NRC
  - Visiting Scholar: University of Copenhagen

Impact Team in Georgian Partners

- Expertise in machine learning, security, privacy, natural language processing, software engineering.
- Enable portfolio companies to accelerate understanding and adoption of thesis areas.
- Act as an extension to a portfolio company's R&D capabilities.
- Engagements range from strategy workshops through to applied research.



From Conventional Machine Learning to Deep Neural Networks

Although it is often valuable to approach problems with conventional machine learning,

✓ Easier to **implement**, **maintain** and **explain**

Current trends in the Machine Learning research community focus on deep neural networks.

What is **deep learning**?

> *Deep Neural networks are stacked, multilayer networks where each layer provides nonlinear information processing and corresponds to a different level of abstraction.*

What are the advantages of DL over conventional Machine Learning?

- More **flexible** to capture rich and intricate patterns of data
- Easier **Transfer Learning** and better **Modularity**
- Automatically learn **high-level feature representations**

Learning Representation

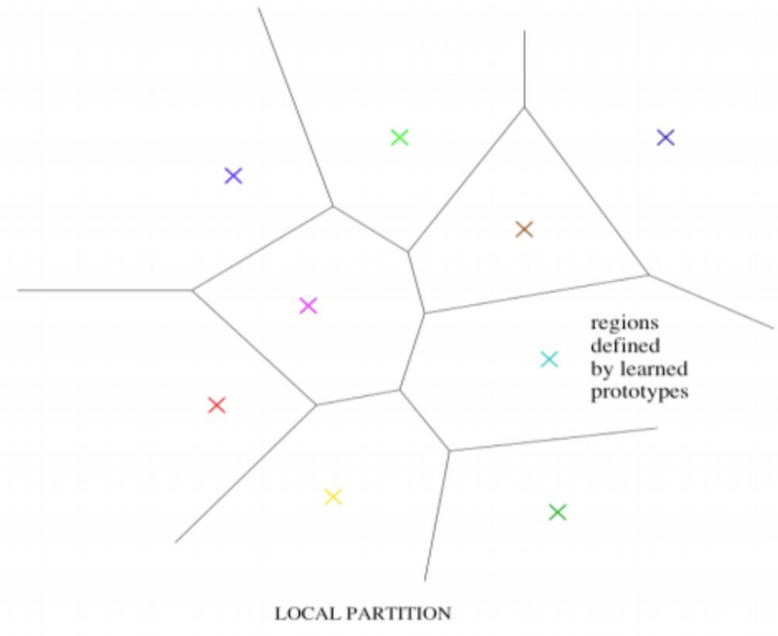
- Handcrafting features:
  - Are Domain/task dependent
  - Requires domain/language expertise
  - Might be over-specified or incomplete
- Why using deep learning to learn representation?
  - Automatic feature learning
  - Learning **distributed representation**
  - Learning **different-level of abstraction**

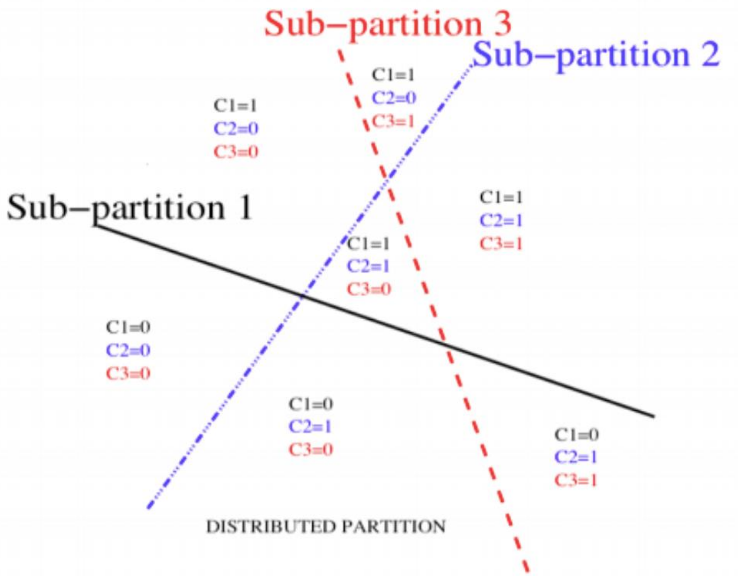
One-hot Encoding Vs Distributed Representation

- One-Hot Encoding:

motel [0 0 0 0 0 0 0 0 0 0 1 0 0 0 0] AND
hotel [0 0 0 0 0 0 0 1 0 0 0 0 0 0 0] = 0

- Distributed Representation



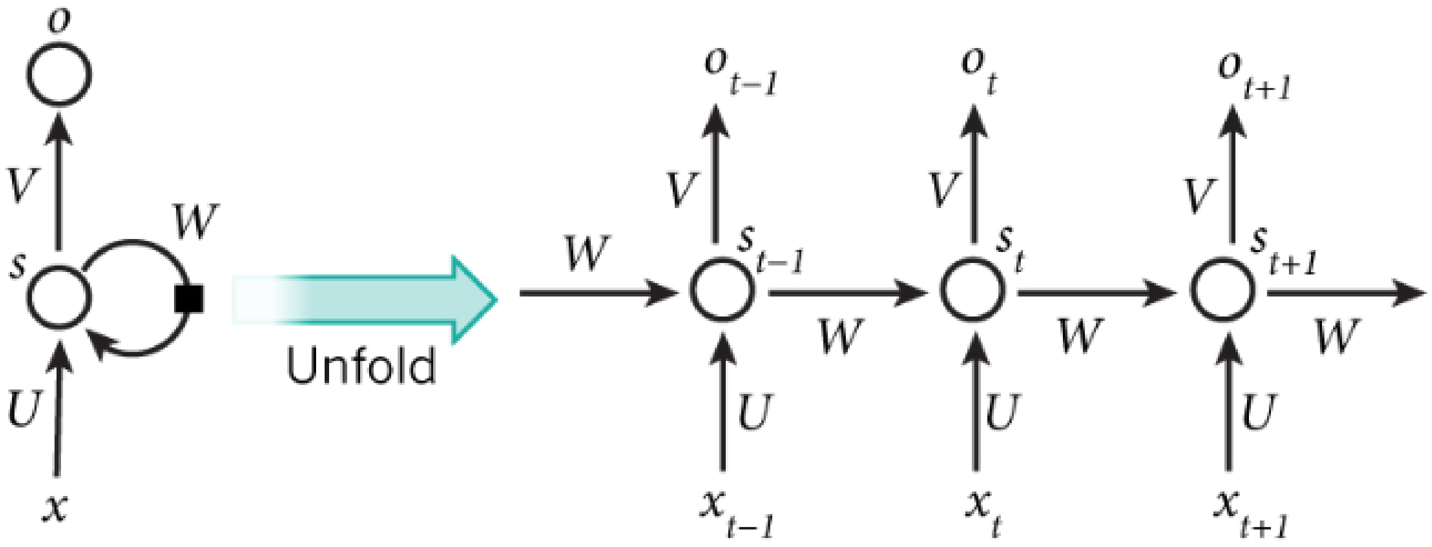


*Figures from Richard Socher Slide

Why Neural Word Embedding?

- Soft clustering models, such as LSA or LDA, learn for each cluster/topic a distribution over words of how likely that word is in each cluster
- Advantages of the neural word embedding approach:
  - We can easily add supervision from one or many tasks to learn task-specific representation
  - We can build representations for large linguistic units

Recurrent Neural Networks (RNN)



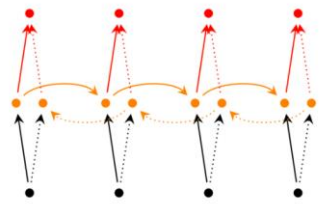
$$s_t = f(Ux_t + Ws_{t-1})$$

$$o_t = \mathrm{softmax}(Vs_t)$$

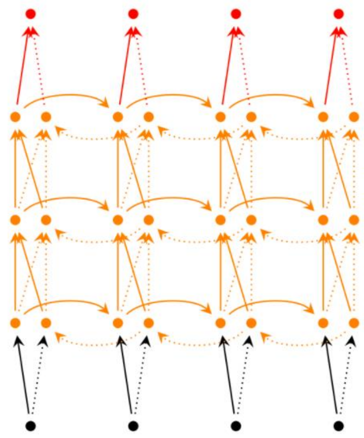
Figure from **WILDML**

RNN Extensions

- **Bidirectional RNNs**



- **Deep (Bidirectional) RNNs**



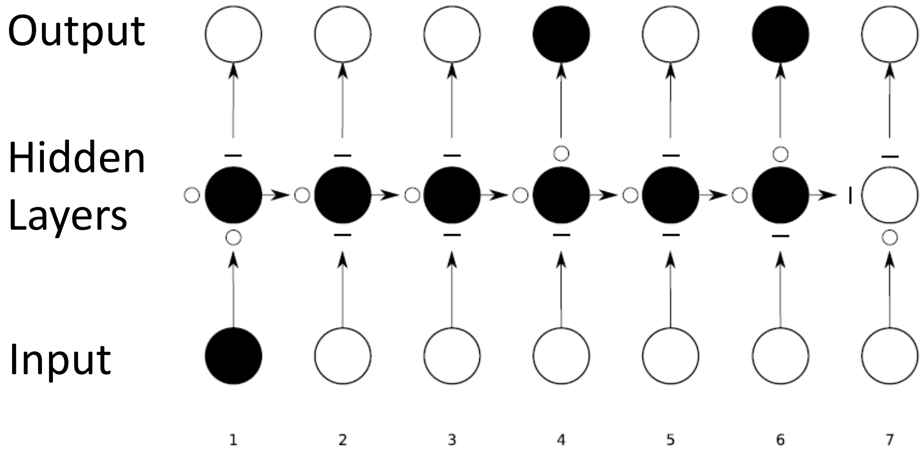
Figures from **WILDML**

Long Short-Term Memory (LSTM)

- LSTM [Hochreiter & Schmihuber ’97] has showed to be effective in a wide range problems:
  - handwriting recognition [Graves, ’08],
  - machine translation [Sutskever, ’14; Cho, ’14],
  - speech recognition [Graves, ’13],
  - image-to-text conversion [Vinyals, ’14],
  - robot control [Mayer, ’08],
  - etc.

Long Short-Term Memory (LSTM)

- In conventional LSTM, history is summarized and encoded in *memory cells* in a **sequential** fashion.



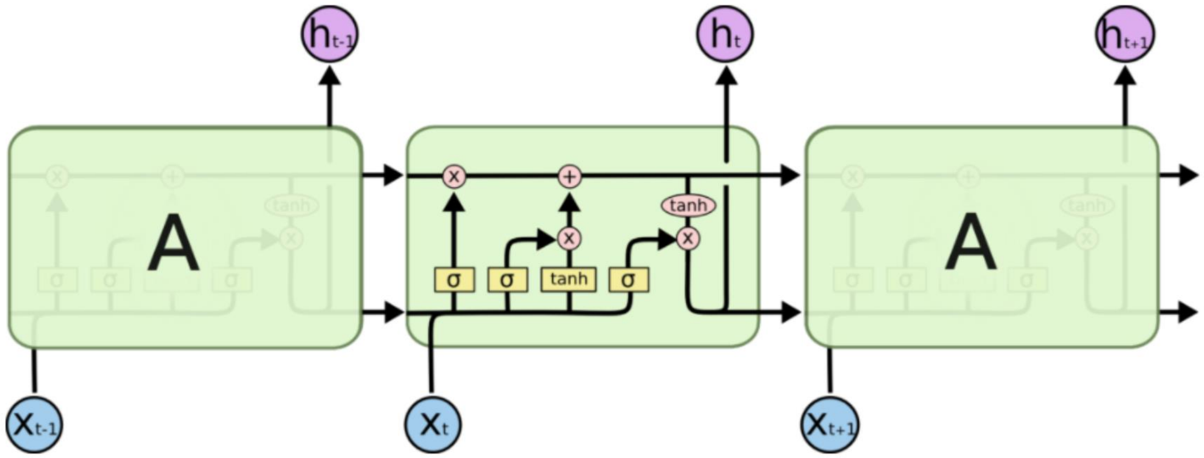
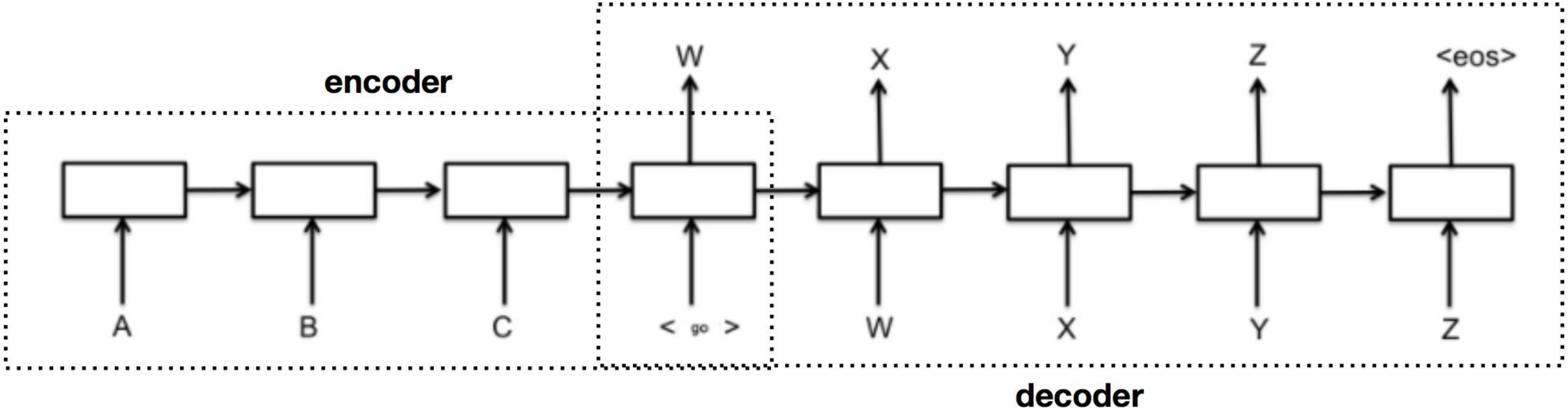
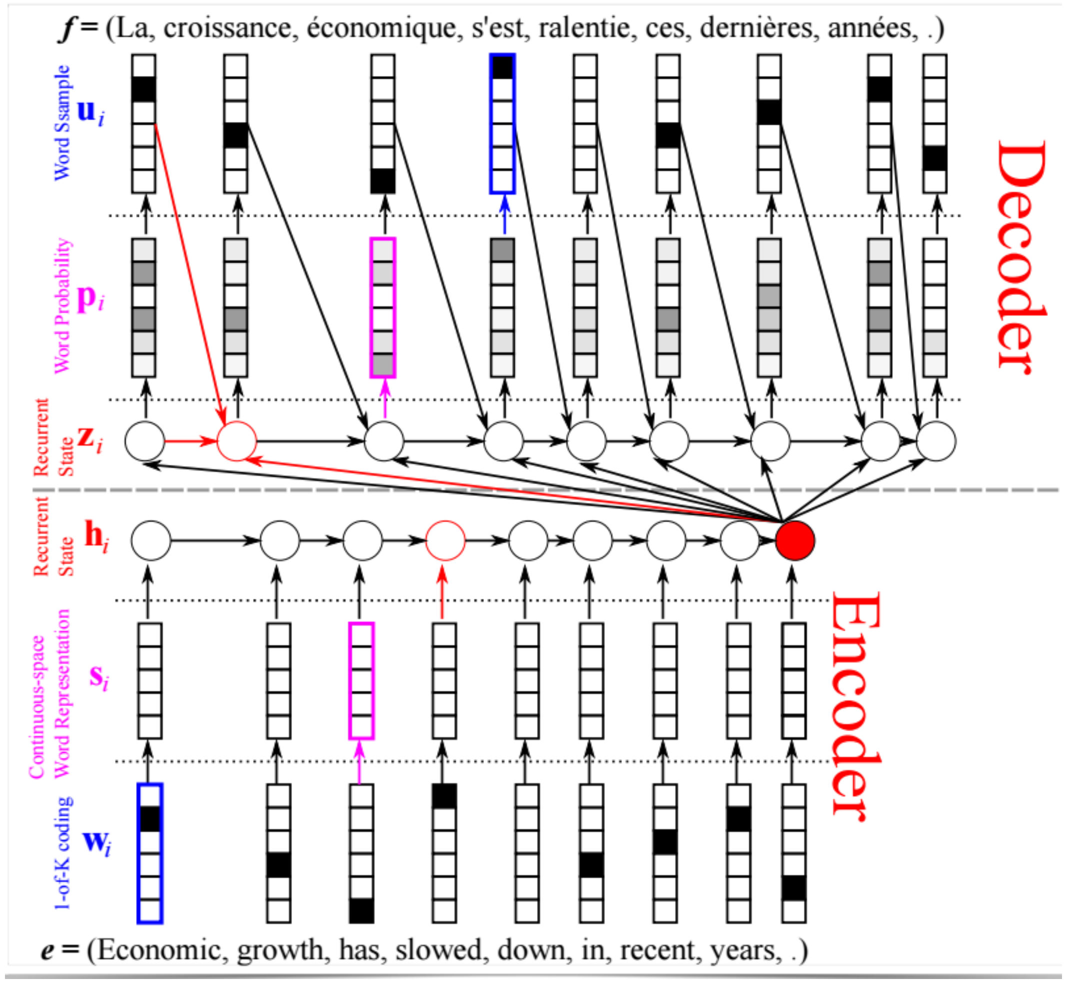


Figure from **Christopher Olah**

Sequence-To-Sequence Model
[Sutskever et al. 2014, Cho et al. 2014]

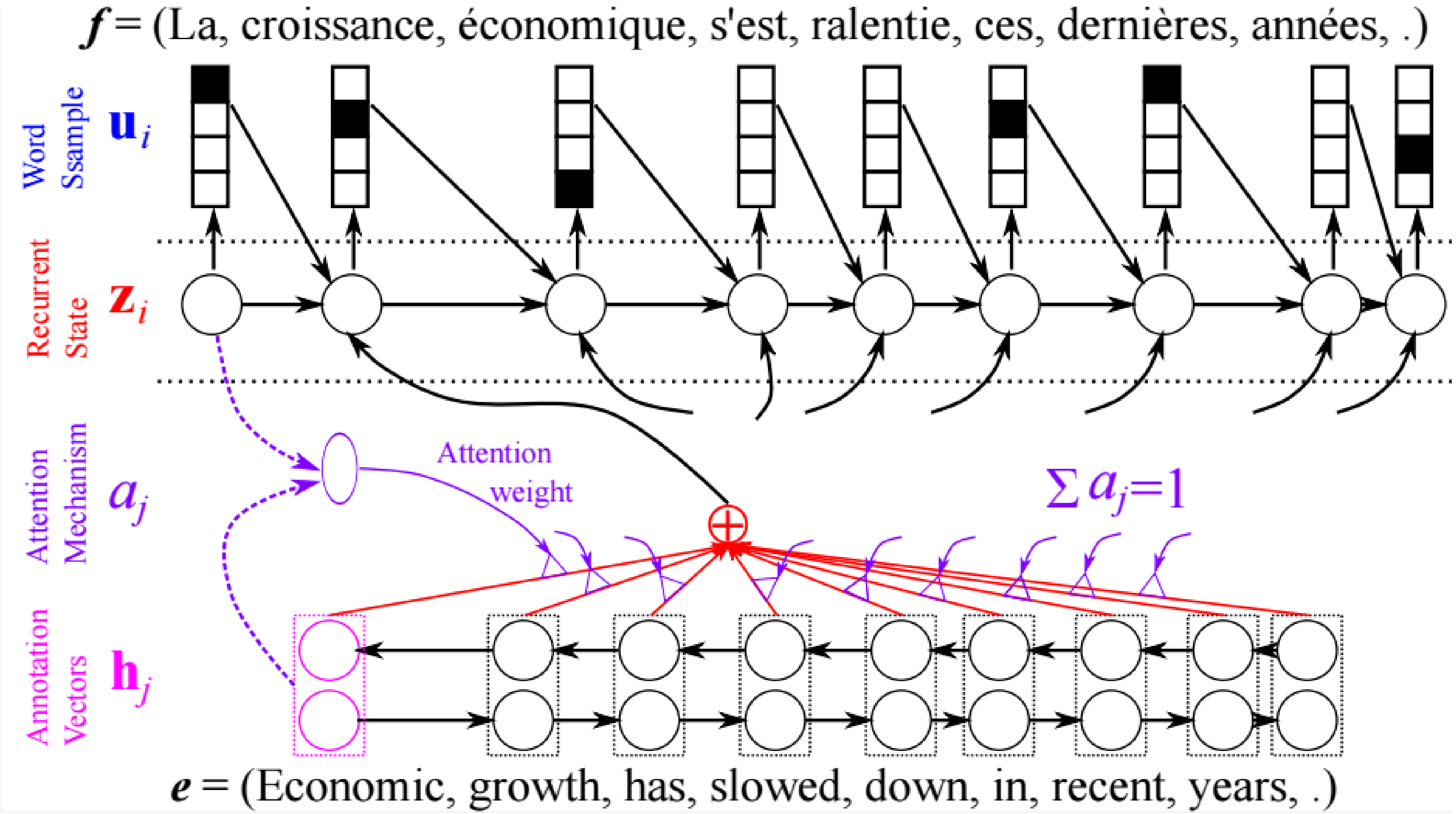


Sequence-To-Sequence Model [Sutskever et al. 2014, Cho et al. 2014]



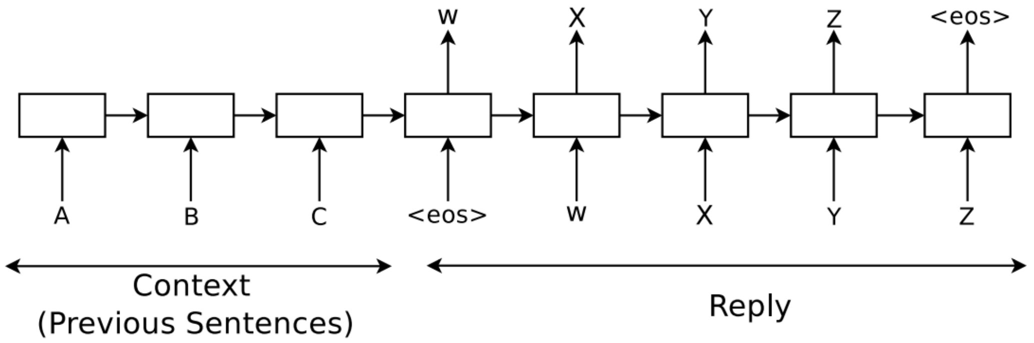
Attention-based Model

[Bahdanau et al. 2015, Luong et al. 2015]



Cho, Montreal Summer School

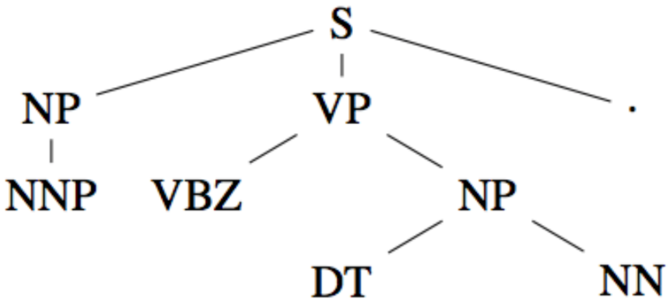
Machine: *hi*
Human: *hello*
Machine: *could you please let me know what are the operating systems you are using ?*
Human: *i am using linux*
Machine: *ok it was in the ssl vpn on which you stated it was broken*
Machine: *please check if you are logged in to corp network*
Machine: *what is the error that you are running please*
Human: *i am seeing an error related to vpn*
Machine: *what is the error message that you are getting when connecting to vpn using network connect ?*
Human: *connection refused or something like that*
Machine: *may i know the version of network connect you connect ?*
Human: *i am not sure i know that*
Machine: *is the network connect prompting for any user certificate*
Machine: *?*
Human: *yes*



Question Answering and Dialogue Generation

Vinyals et al. 2015

John has a dog . →



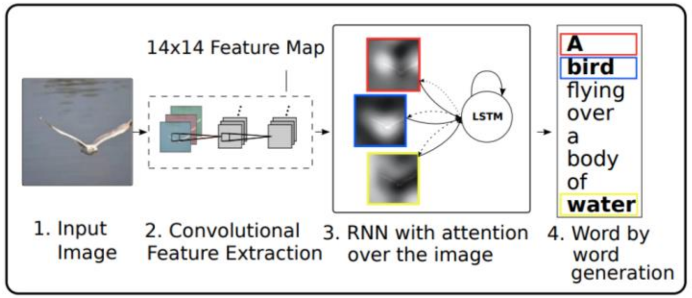
John has a dog . → (S (NP NNP $)_{NP}$ (VP VBZ (NP DT NN $)_{NP}$ $)_{VP}$. $)_{S}$

Grammar as a foreign language

Syntactic constituency parsing

Vinyals et al. 2015

Image Caption Generation (Xu et al, 2015)

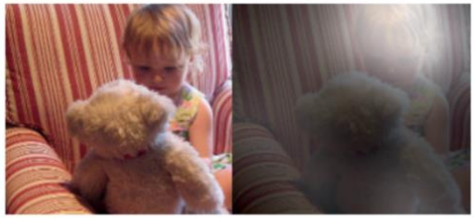


- Images are just another language!

A woman is throwing a **frisbee** in a park.

A **dog** is standing on a hardwood floor.

A **stop** sign is on a road with a mountain in the background



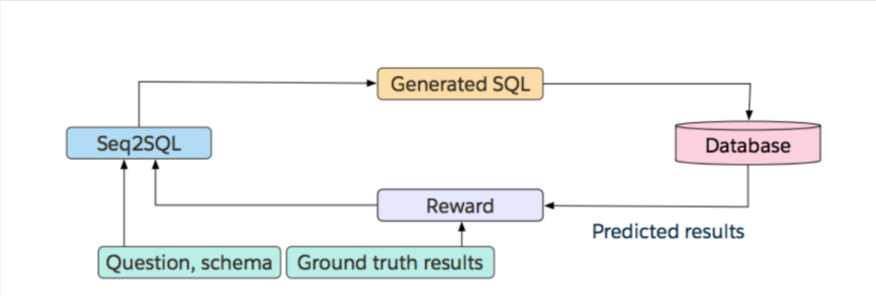
A little **girl** sitting on a bed with a teddy bear.

A group of **people** sitting on a boat in the water.

A giraffe standing in a forest with **trees** in the background.

SQL Queries Generation (Zhong et al, 2017)

- Using policy-based reinforcement learning to generate the conditions of the query, which are unsuitable for optimization using cross entropy loss
- Seq2SQL leverages the structure of SQL to prune the space of generated queries and significantly simplify the generation problem



Q	how many voters from the bronx voted for the socialist party?
P	`SELECT MIN % party = socialist`
S'	`SELECT COUNT the bronx where the bronx = socialist`
S	`SELECT COUNT the bronx WHERE the bronx = socialist`
G	`SELECT the bronx WHERE party = socialist`

Neural-storyteller (Kiros et al, 2015)

Generated story about image
Model: Romantic Novels

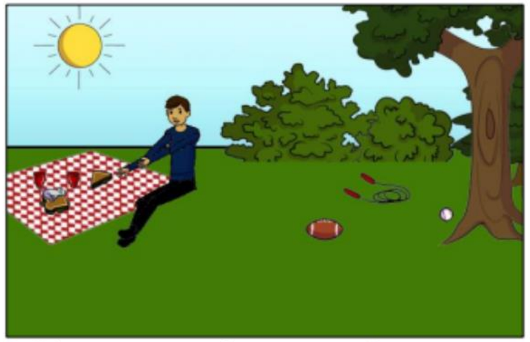
"He was a shirtless man in the back of his mind, and I let out a curse as he leaned over to kiss me on the shoulder.

He wanted to strangle me, considering the be-atiful boy I'd become wearing his boxers."

Visual Question Answering (Agrawal et al, 2016)

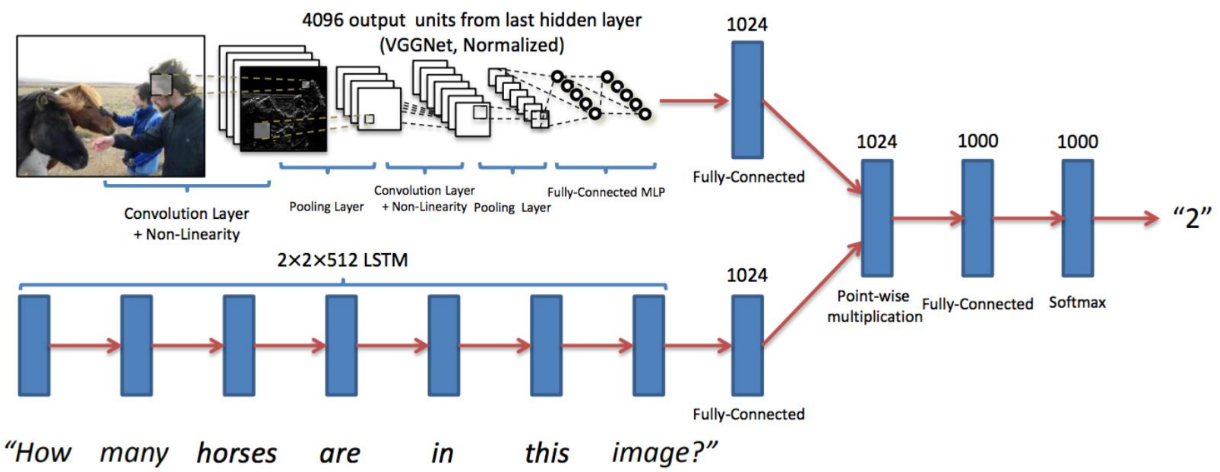
What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

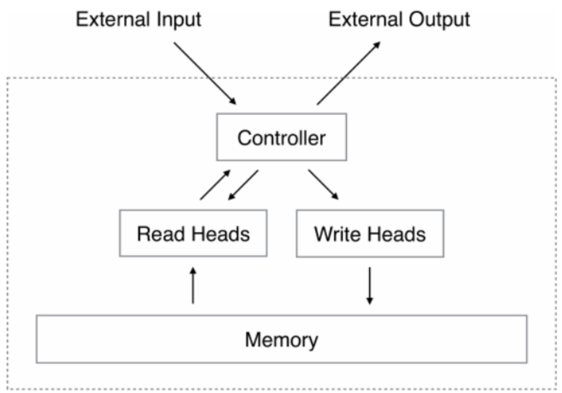


Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?



Other Approaches



- Pointer Networks (Vinyals et al, 2015)

  It differs from the previous attention attempts in that, instead of using attention to blend hidden units of an encoder to a context vector at each decoder step, it uses attention as a pointer to select a member of the input sequence as the output.

- Neural Turing Machines **(**Graves et al. 2014**)**

  A Neural Turing Machine (NTM) architecture contains two basic components: a neural network controller and a memory bank. Like most neural networks, the controller interacts with the external world via input and output vectors. Unlike a standard network, it also interacts with a memory matrix using selective read and write operations.

- Memory networks (Sukhbaatar et al, 2015)

Other Resources

- DeeDeep Learning for NLP (without Magic) https://nlp.stanford.edu/courses/NAACL2013/NAACL2013-Socher-Manning-DeepLearning.pdf
- Recurrent Neural Networks Tutorial http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/
- Deep Learning for Machine Translation https://drive.google.com/drive/folders/0B16RwCMQqrtda2toU29Za19qcWM