Text Prediction using Recurrent Neural Networks and TensorFlow

Dipendra Jha and Reda Al-Bahrani

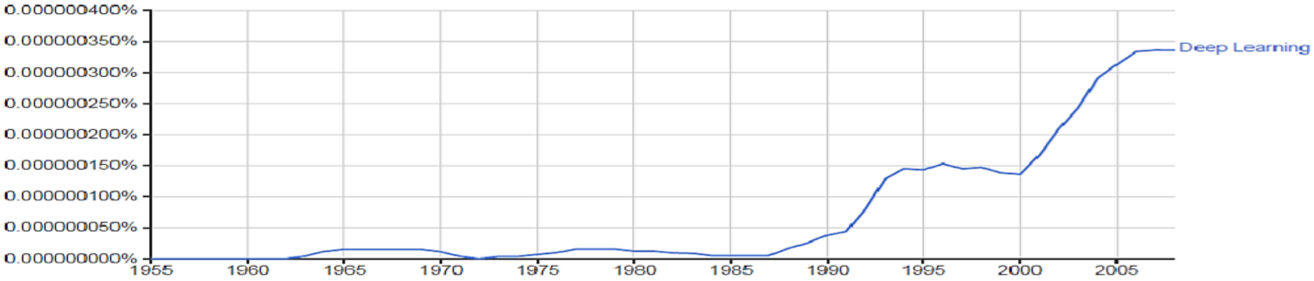
Northwestern University

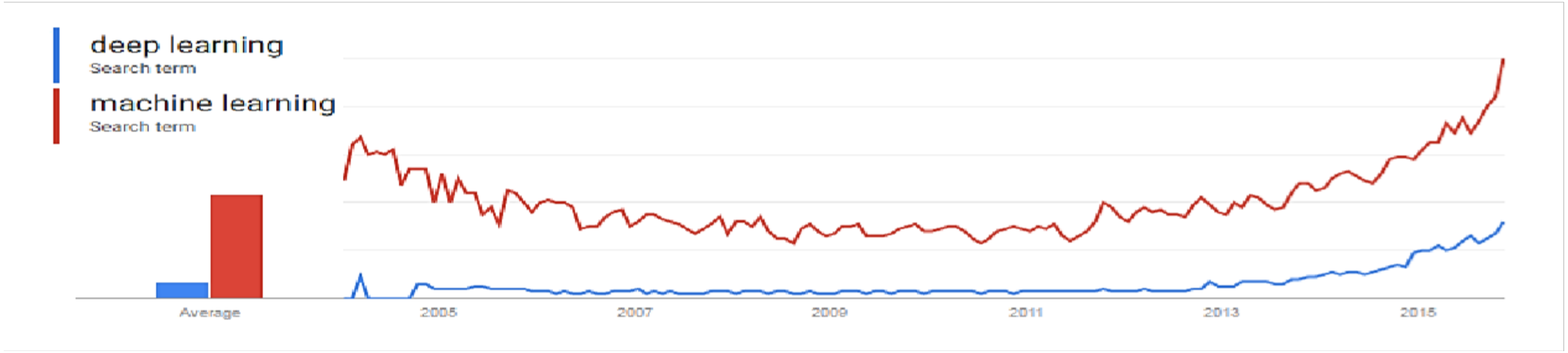
dipendra@u.northwestern.edu, reda@u.northwestern.edu

Outline

- **Deep Learning**
- Text Prediction
- Recurrent Neural Networks (RNN, LSTM)
- Word Embeddings
- Brief Introduction to TensorFlow
- Text Generation using TF
- Training Demonstration using TensorFlow

Deep Learning Interest: Google Trends



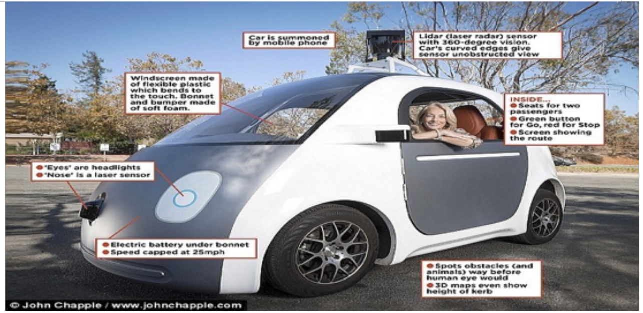


Big Players in Deep Learning

facebook
Microsoft
YAHOO!
Google
IBM
NVIDIA
Bai du 百度

Just with in Google

- Search
- Search by image
- Driverless cars
- Youtube recommendations
  - Videos
  - Thumbnails
- Maps
  - Reading street adresses

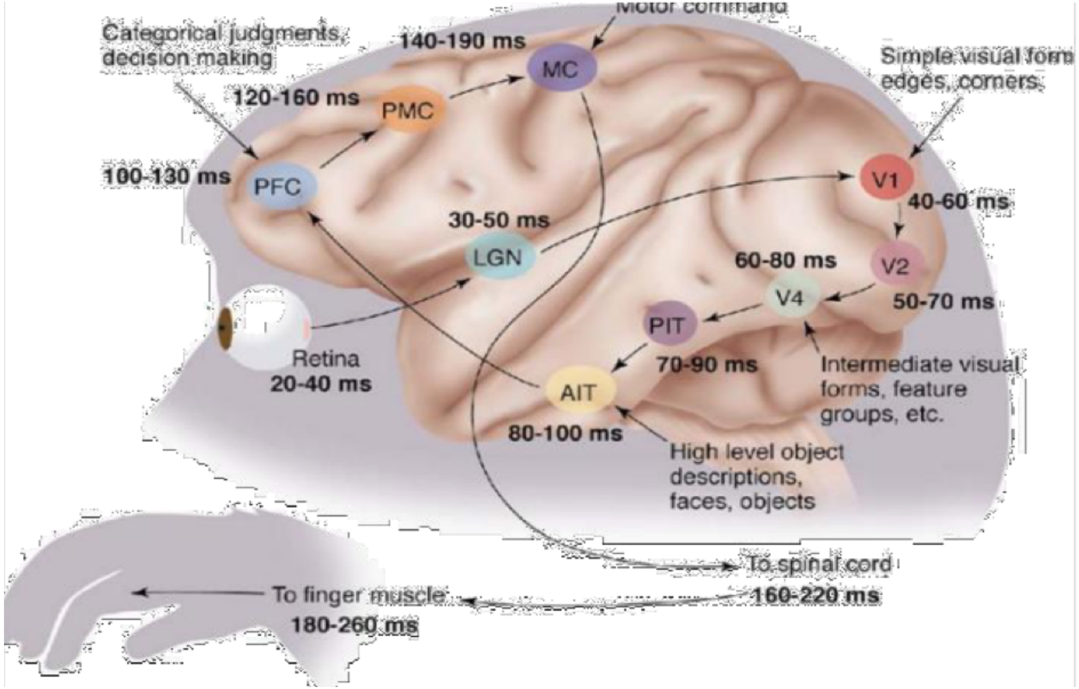

Car is summoned by mobile phone
Lidar (laser radar) sensor with 360-degree vision. Car's curved edges give sensor unobstructed view
Windscreen made of flexible plastic which bends to the touch. Bonnet and bumper made of soft foam.
INSIDE...
● Seats for two passengers
● Green button for Go, red for Stop
● Screen showing the route
● 'Eyes' are headlights
● 'Nose' is a laser sensor
● Electric battery under bonnet
● Speed capped at 25mph
● Spots obstacles (and animals) way before human eye would
● 3D maps even show height of kerb

© John Chapple / www.johnchapple.com

What is Deep Learning?

- Part of machine learning field of learning representation of data. Exceptionally effective at learning patterns.
- Learning algorithms that derive meaning out of data by using a hierarchy of multiple layers that mimic the neural networks of our brain.
- If you provide the system tons of information, it begins to understand it and respond in useful ways.

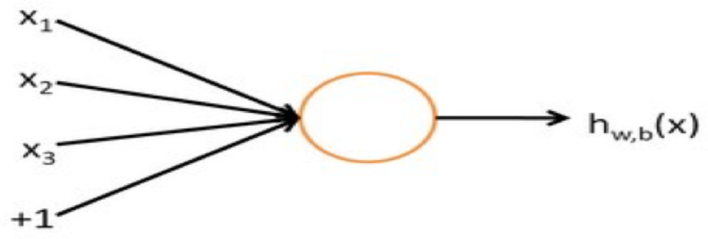
Inspired by Brain



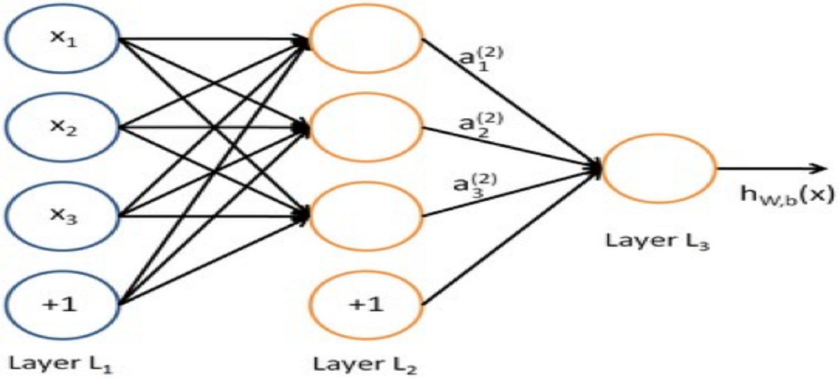
The first **hierarcy of neurons** that receives information in the visual cortex are sensitive to specific edges while brain regions further down the visual pipeline are sensitive to more complex structures such as faces.

(Artificial) Neural Networks and Training

A ‘neuron’



A small neural network ‘neuron’



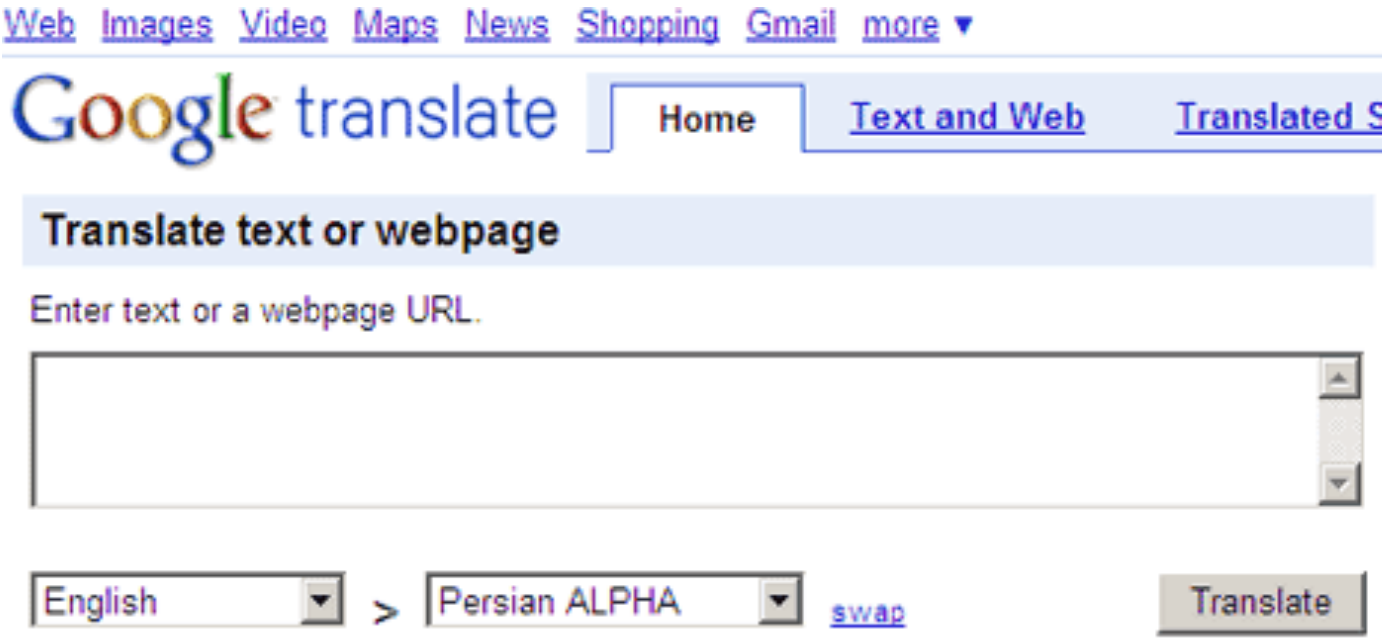
How to define Deep Learning?

- Deep learning: the most cutting edge ML/AI research
- "A method which makes predictions by using a sequence of non-linear processing stages. The resulting intermediate representations can be interpreted as feature hierarchies and the whole system is jointly learned from data." - Facebook Research
- "Machine that learn to represent the world."
- "End-to-end machine learning." - Yann LeCun

Outline

- **Deep Learning**
- **Text Prediction**
- Recurrent Neural Networks (RNN, LSTM)
- Word Embeddings
- Brief Introduction to TensorFlow
- Text Generation using TF
- Training Demonstration using TensorFlow

Why Text Prediction?



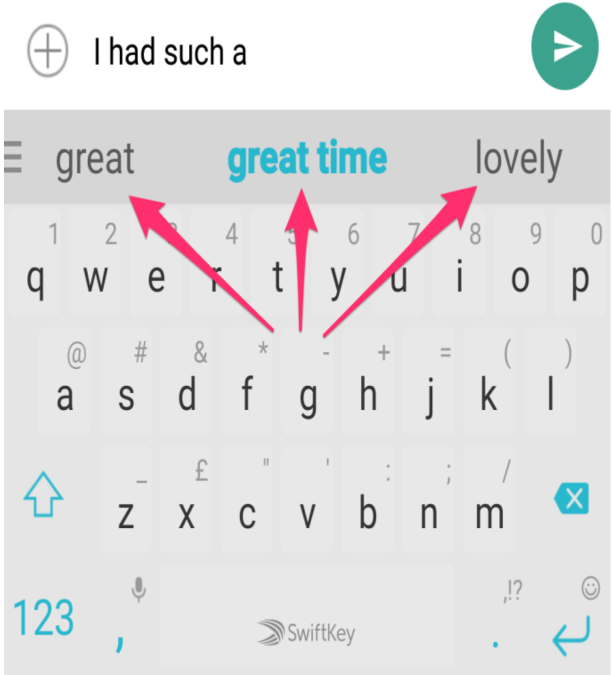
Translate text or webpage

Enter text or a webpage URL.

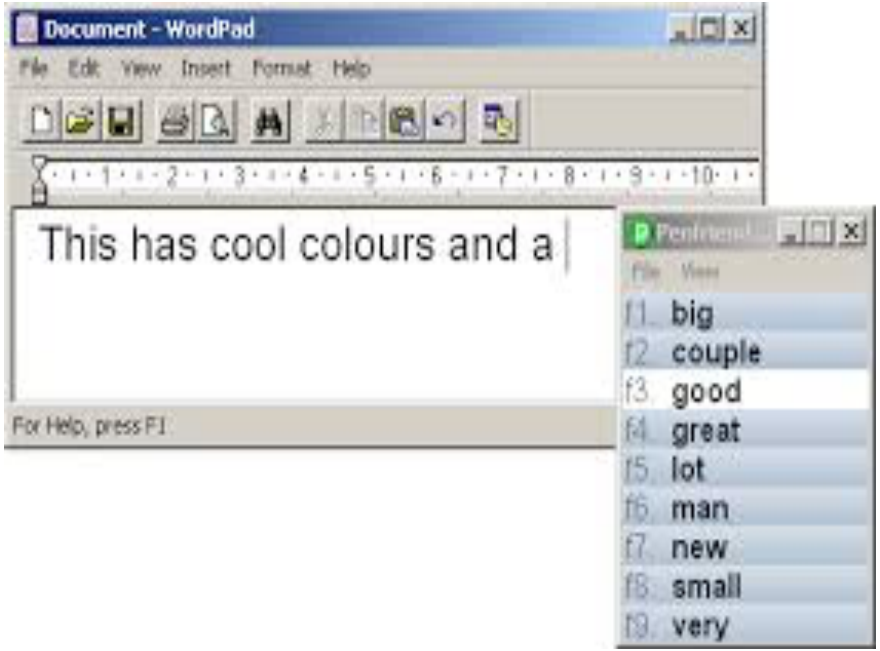
English > Persian ALPHA

Why Text Prediction?

Why Text Prediction?

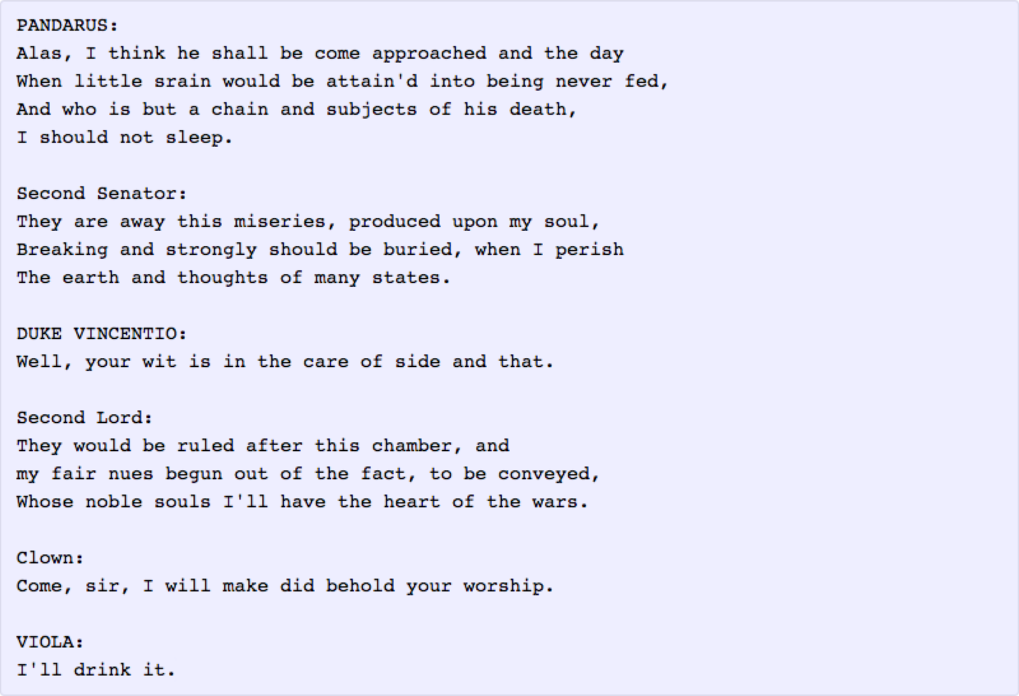


I had such a



This has cool colours and a

Why Text Prediction?

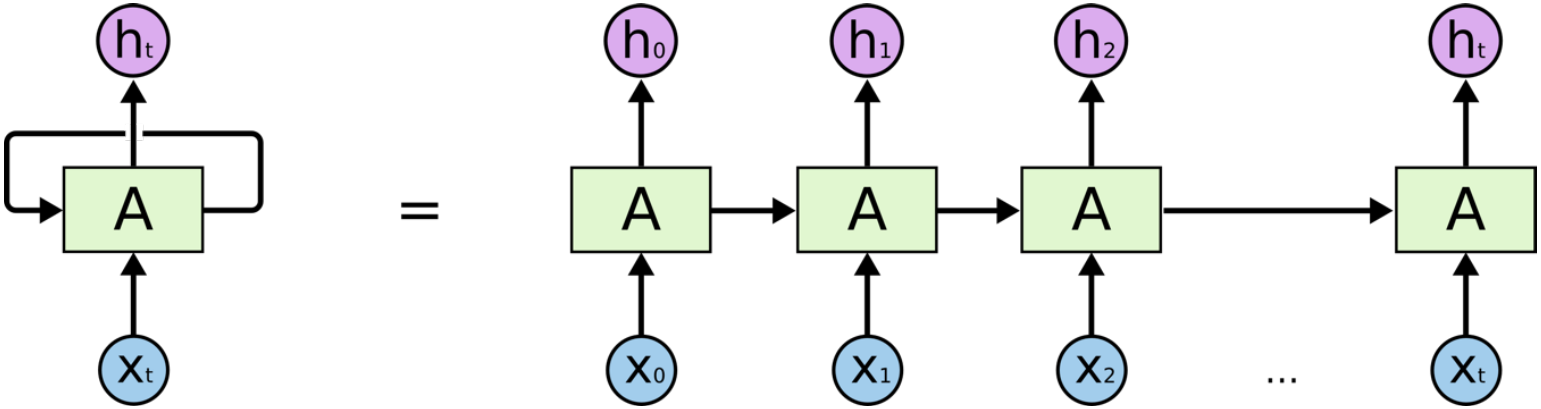


```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

Outline

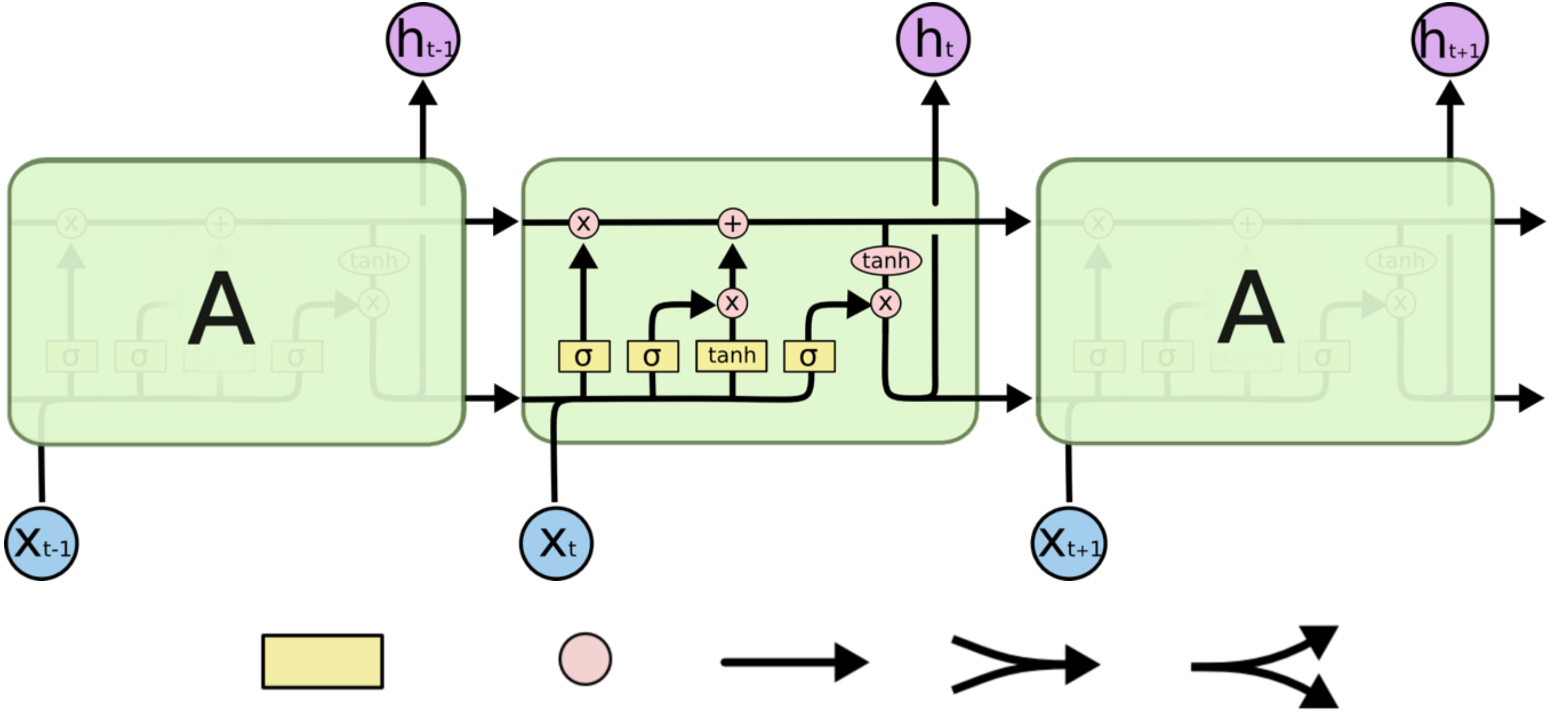
- **Deep Learning**
- **Text Prediction**
- **Recurrent Neural Networks (RNN, LSTM)**
- Word Embeddings
- Brief Introduction to TensorFlow
- Text Generation using TF
- Training Demonstration using TensorFlow

Recurrent Neural Networks



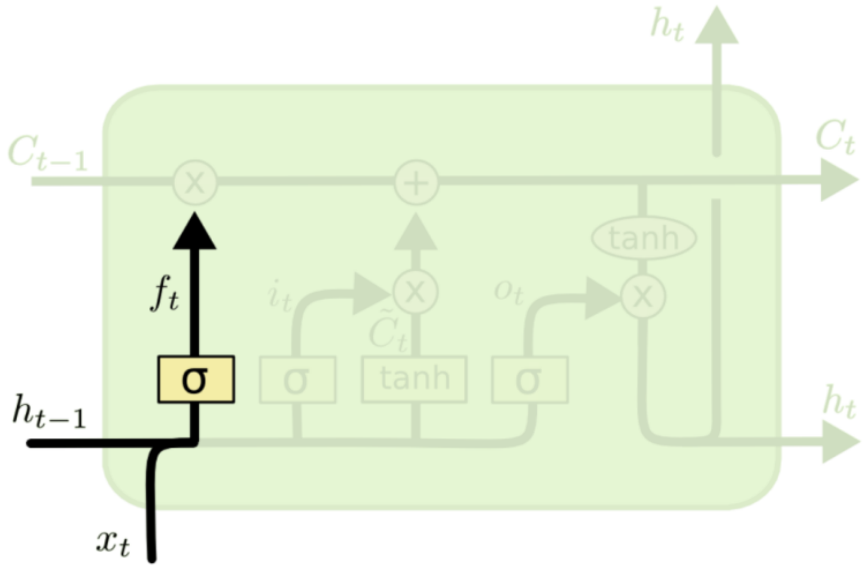
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long short Term Memory



http://colah.github.io/posts/2015-08-Understanding-LSTMs/

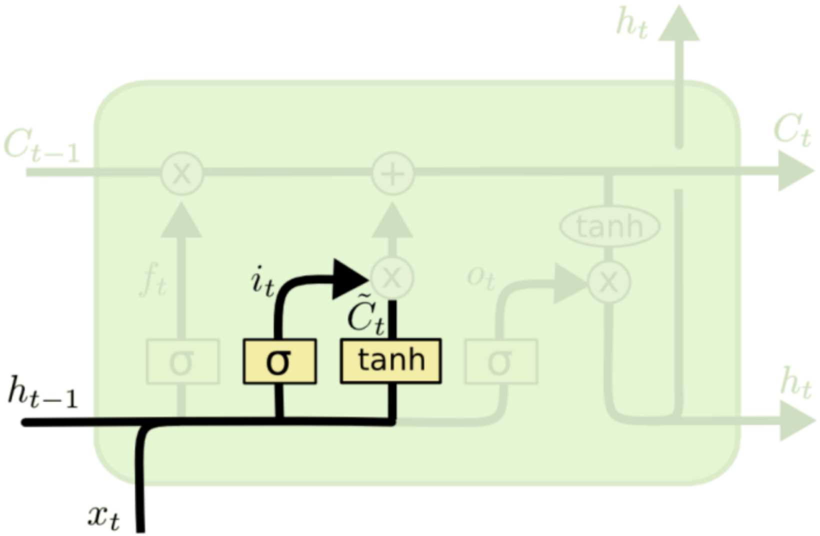
Long short Term Memory (formulas)



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long short Term Memory (formulas)

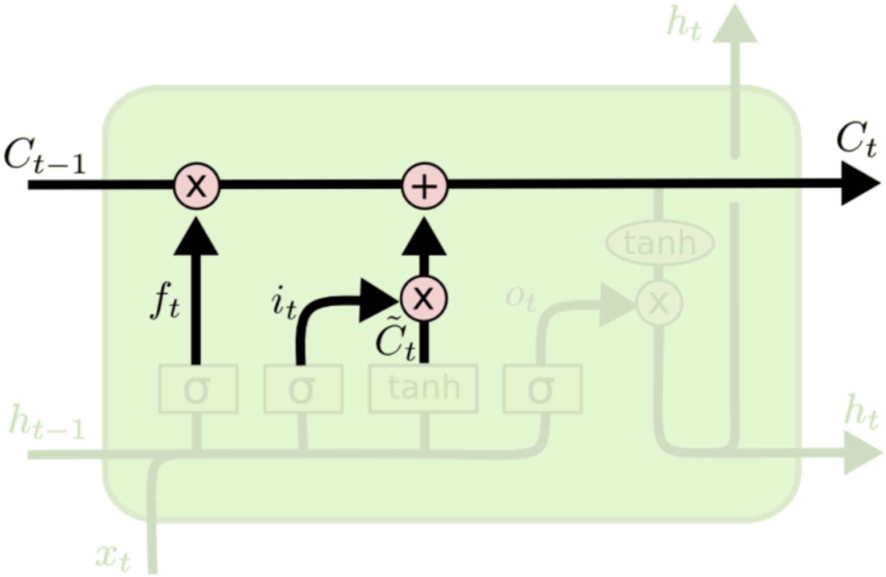


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

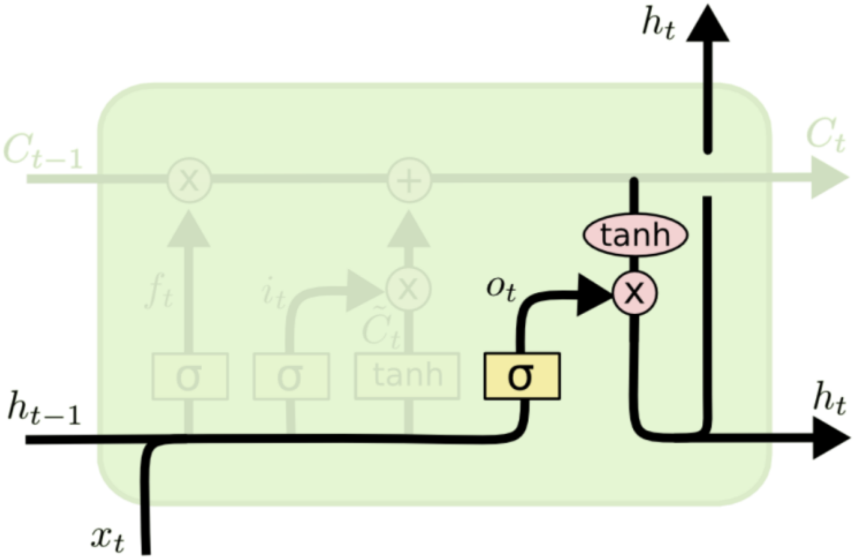
Long short Term Memory (formulas)



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long short Term Memory (formulas)



$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Outline

- **Deep Learning**
- **Text Prediction**
- **Recurrent Neural Networks (RNN, LSTM)**
- **Word Embeddings**
- Brief Introduction to TensorFlow
- Text Generation using TF
- Training Demonstration using TensorFlow

Word Embeddings

- Word2Vec
- CBOW
- Skip-gram

http://sebastianruder.com/word-embeddings-1/index.html

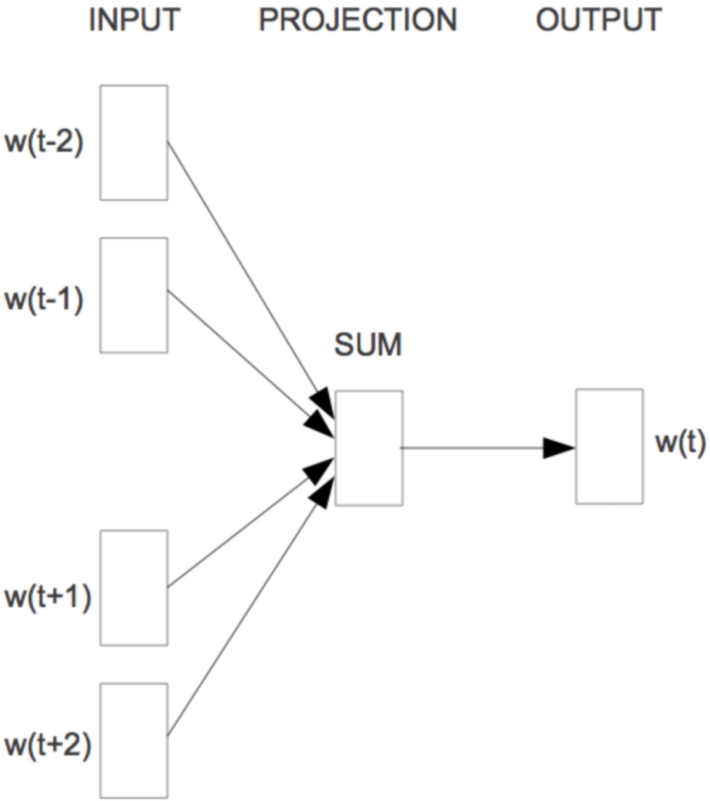
Word2Vec

- word2vec is not deep learning.
- No expensive hidden layers.
- Somehow take additional context into account.

king−man+woman≈queen

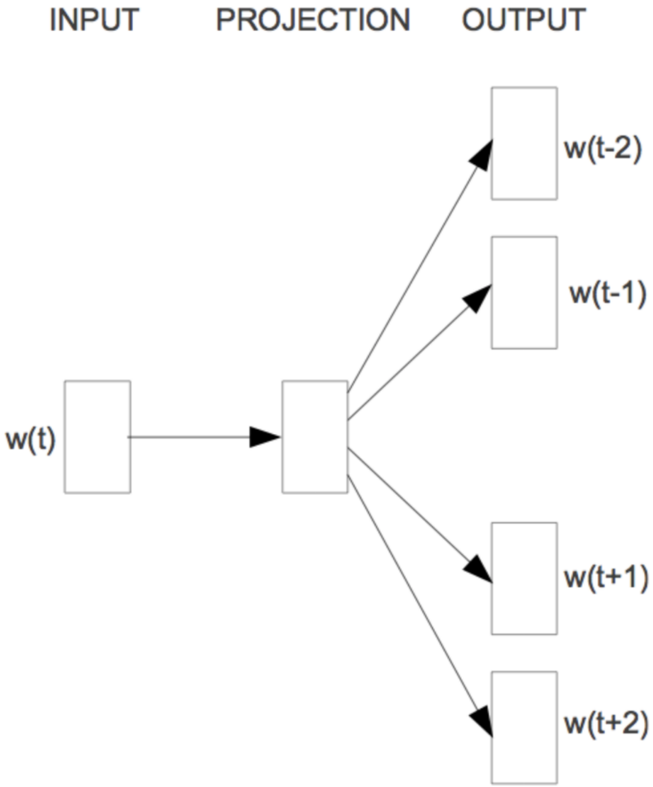
Continuous bag-of-words (CBOW)

- Language models look at the past words for its predictions.
- Words before and after the target word to predict.
- Order is of no importance.



Skip-gram

- Instead of using the surrounding words to predi the centre word as with CBOW.
- Skip-gram uses the centre word to predict the surrounding words.
- The skip-gram architecture does not contain a hidden layer.



Outline

- **Deep Learning**
- **Text Prediction**
- **Recurrent Neural Networks (RNN, LSTM)**
- **Word Embeddings**
- **Brief Introduction to TensorFlow**
- Text Generation using TF
- Training Demonstration using TensorFlow

TensorFlow

- A machine learning library created by Google.
- Started as a research project.
- Generates a computational graph like Theano.
- Everything about TensorFlow is here:
  - https://www.tensorflow.org

TensorFlow Recipe

- Recipe for a TensorFlow application:
  - Define a series of expressions
  - Initialize variables
  - Start a session (launch a graph)
  - Run the graph, feed some data, fetch some values

TensorFlow Essentials

- Four types of objects make TensorFlow unique from other frameworks
  - Session
  - Computational graph
  - Variables
  - Placeholder

Outline

- **Deep Learning**
- **Text Prediction**
- **Recurrent Neural Networks (RNN, LSTM)**
- **Word Embeddings**
- **Brief Introduction to TensorFlow**
- **Text Generation using TF**
- Training Demonstration using TensorFlow

Steps in Text Generation

1. Clean Data
2. Build Vocabulary
3. Convert text to Word Vectors
4. Define the Model (Encoder-Decoder)
5. Train the model
6. Generate text

Cleaning Data

- Tokenization
- Lemattization
- Stemming

Converting text to Word Vectors

- A vocabulary of all the words in input data
- An unique mapping from each word to an index
- Use the word mapping to convert the input text to Word Vectors

Defining a model

1. Use a word embedding to embed the input word into 2D arrays.
2. Use a decoder (with encoder) composing of RNNs or LSTMs to make the predictions.
3. Apply some fully connected layers on top of the decoder output to make the actual predictions.

Training the model and Generating the Text

- Start a session.
- Initialize variables.
- Train the model for some number of epochs.
- Make the actual text predictions.

Outline

- **Deep Learning**
- **Text Prediction**
- **Recurrent Neural Networks (RNN, LSTM)**
- **Word Embeddings**
- **Brief Introduction to TensorFlow**
- **Text Generation using TF**
- **Training Demonstration using TensorFlow**

Training Demonstration

http://localhost:8888/notebooks/TensorFlow/TextGeneration.ipynb

Conclusion

- Deep Learning is no magic!
- Learnt about how RNNs and LSTMs work.
- Learnt about generating text using LSTMs and TF.