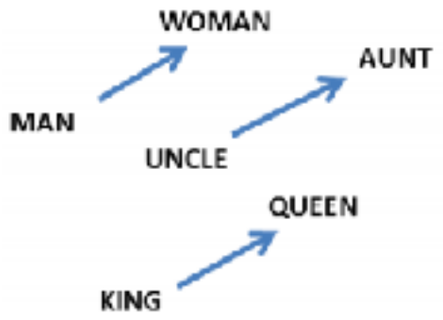
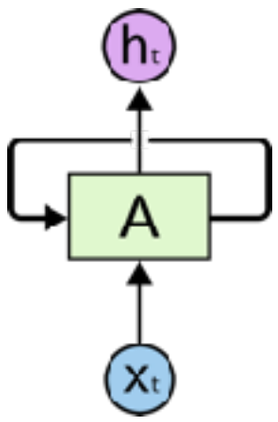
Word Embeddings and Recurrent Neural Networks

A **Word Embedding** is a representation of words where each word is expressed as a real-valued vector. Words with similar meanings have similar vectors.



[2]

A **Recurrent Neural Network** (RNN) is a class of artificial neural network where connections between units form a directed cycle. This allows it to exhibit dynamic temporal behavior. [1]



[3]

[1] https://en.wikipedia.org/wiki/Recurrent_neural_network
[2] Mikolov et al, 2013
[3] http://colah.github.io/posts/2015-08-Understanding-LSTMs

Word Embeddings

Influences and history

You shall know a word by the company it keeps [1]

LSA (Latent Semantic Analysis) "boat" related to "water" [2]

LDA (Latent Dirichlet Allocation) cat: milk, meow, kitten dog: puppy, bark, bone [3]

Use documents as context
count-based

[4] Neural language models "boat" related to "ship"

word2vec [5]

Use words as context
predictive

[1] https://en.wikipedia.org/wiki/John_Rupert_Firth
[2] https://en.wikipedia.org/wiki/Latent_semantic_analysis
[3] https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation
[4] Bengio et al., 2003
[5] Mikolov et al., 2013
https://en.wikipedia.org/wiki/Word_embedding

Word Embeddings

Map words to vectors

$$W : words \rightarrow \mathbb{R}^n$$

$$\text{W("cat")} = (0.2, -0.4, 0.7, ...)$$
$$\text{W("mat")} = (0.1, -0.6, -0.1, ...)$$

Tasks

Similarities between terms
Classification
Document clustering
Named Entity recognition
Sentiment analysis

http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/
https://www.tensorflow.org/tutorials/word2vec

Word Embeddings

Similarities between terms

FRANCE	JESUS	XBOX	REDDISH	SCRATCHED	MEGABITS
454	1973	6909	11724	29869	87025
AUSTRIA	GOD	AMIGA	GREENISH	NAILED	OCTETS
BELGIUM	SATI	PLAYSTATION	BLUISH	SMASHED	MB/S
GERMANY	CHRIST	MSX	PINKISH	PUNCHED	BIT/S
ITALY	SATAN	IPOD	PURPLISH	POPPED	BAUD
GREECE	KALI	SEGA	BROWNISH	CRIMPED	CARATS
SWEDEN	INDRA	PSNUMBER	GREYISH	SCRAPED	KBIT/S
NORWAY	VISHNU	HD	GRAYISH	SCREWED	MEGAHERTZ
EUROPE	ANANDA	DREAMCAST	WHITISH	SECTIONED	MEGAPIXELS
HUNGARY	PARVATI	GEFORCE	SILVERY	SLASHED	GBIT/S
SWITZERLAND	GRACE	CAPCOM	YELLOWISH	RIPPED	AMPERES

Collobert, Ronan, et al., 2011

Word Embeddings

word2vec

the quick brown fox jumped over the lazy dog

window size: 1
context: left word, right word

Continuous Bag of Words (CBOW)
predict target word from context
smoother because uses whole context
useful for smaller datasets

[the, brown], quick
[quick, fox], brown
[brown, jumped], fox
…

Skip-gram
predict context from target word
better with larger datasets

quick, the
quick, brown
brown, quick
brown, fox
fox, brown
fox, jumped
…

Mikolov et al., 2013
https://en.wikipedia.org/wiki/Word2vec
https://www.tensorflow.org/tutorials/word2vec

Word Embeddings

word2vec

Traditional approach
Use maximum likelihood
expensive

$$\begin{aligned} \mathrm{J}_{\mathrm{ML}} &= \log P(w_t|h) \\ &= \mathrm{score}(w_t, h) - \log\left(\sum_{\text{Word w' in Vocab}} \exp\{\mathrm{score}(w', h)\}\right) \end{aligned}$$

word2vec
binary classifier with negative sampling
more efficient

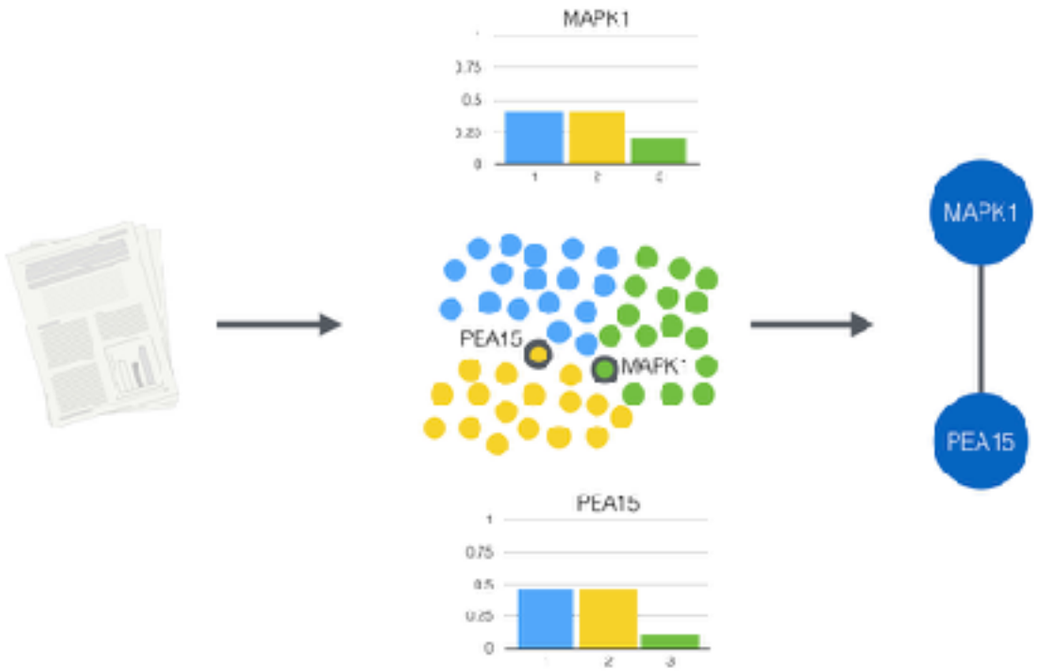
$$J_{\mathrm{NEG}}^{(t)} = \log Q_\theta(D = 1|\text{fox, brown}) + \log(Q_\theta(D = 0|\text{sheep, brown}))$$

https://www.tensorflow.org/tutorials/word2vec

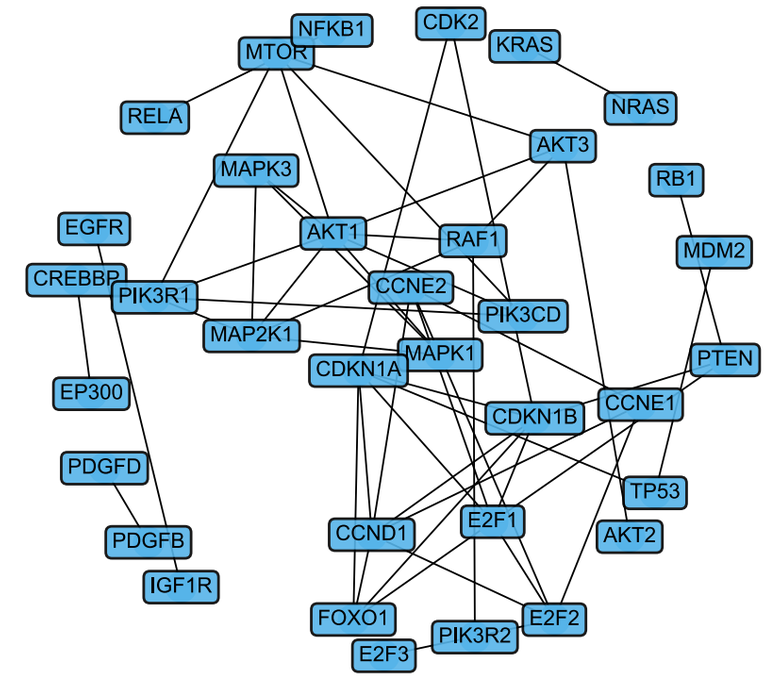
Word Embeddings

INtERAcT

Interaction **N**etwork Inf**E**rence from Vecto**R** Represent**AT**ion of Words

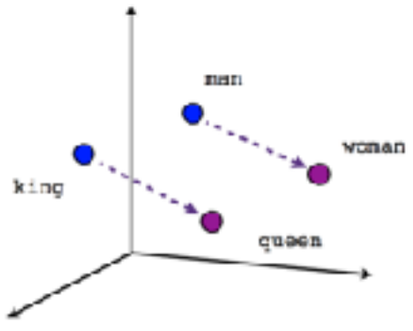


Prostate Cancer (top 50 predicted)



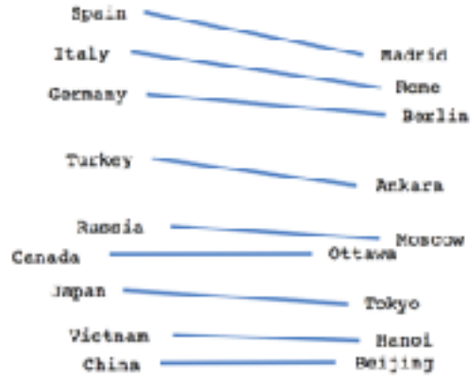
Word Embeddings

Visualization and vector operations



Male-Female

Verb tense



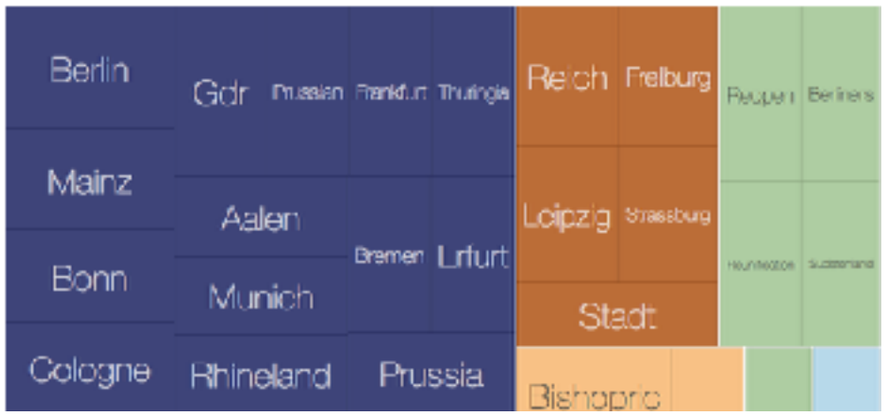
Country-Capital

[1]

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

[2]

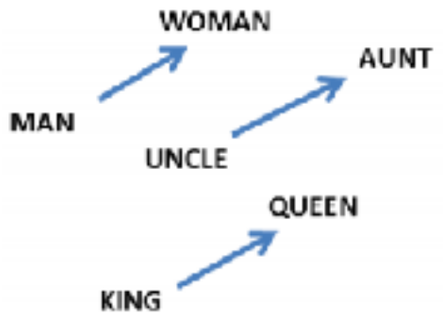
rome -italy +germany (demo)



[1] https://www.tensorflow.org/tutorials/word2vec
[2] Mikolov et al., 2013

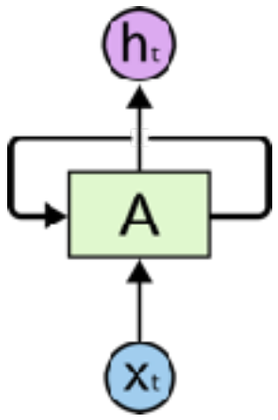
Word Embeddings and Recurrent Neural Networks

A **Word Embedding** is a representation of words where each word is expressed as a real-valued vector. Words with similar meanings have similar vectors.



[2]

A **Recurrent Neural Network** (RNN) is a class of artificial neural network where connections between units form a directed cycle. This allows it to exhibit dynamic temporal behavior. [1]



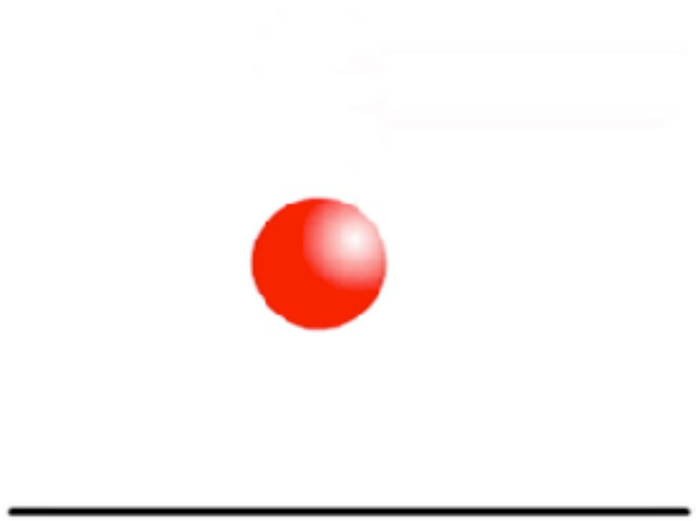
[3]

[1] https://en.wikipedia.org/wiki/Recurrent_neural_network
[2] Mikolov et al, 2013
[3] http://colah.github.io/posts/2015-08-Understanding-LSTMs

Recurrent Neural Networks

Sequences

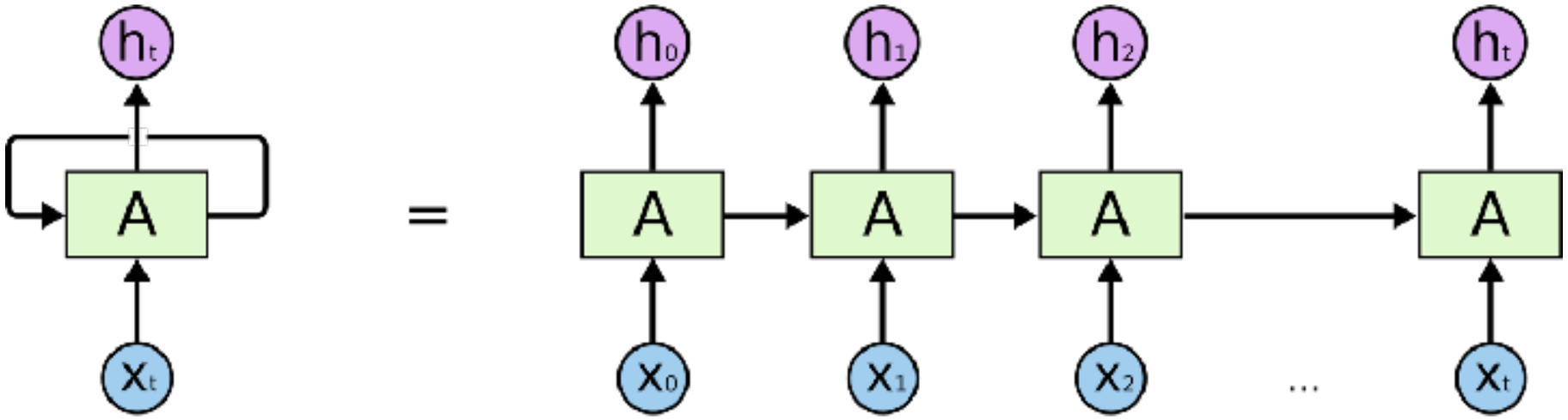
I grew up in France… I speak fluent _____.



http://colah.github.io/posts/2015-08-Understanding-LSTMs/
https://iamtrask.github.io/2015/11/15/anyone-can-code-lstm/

Recurrent Neural Networks

Unrolled

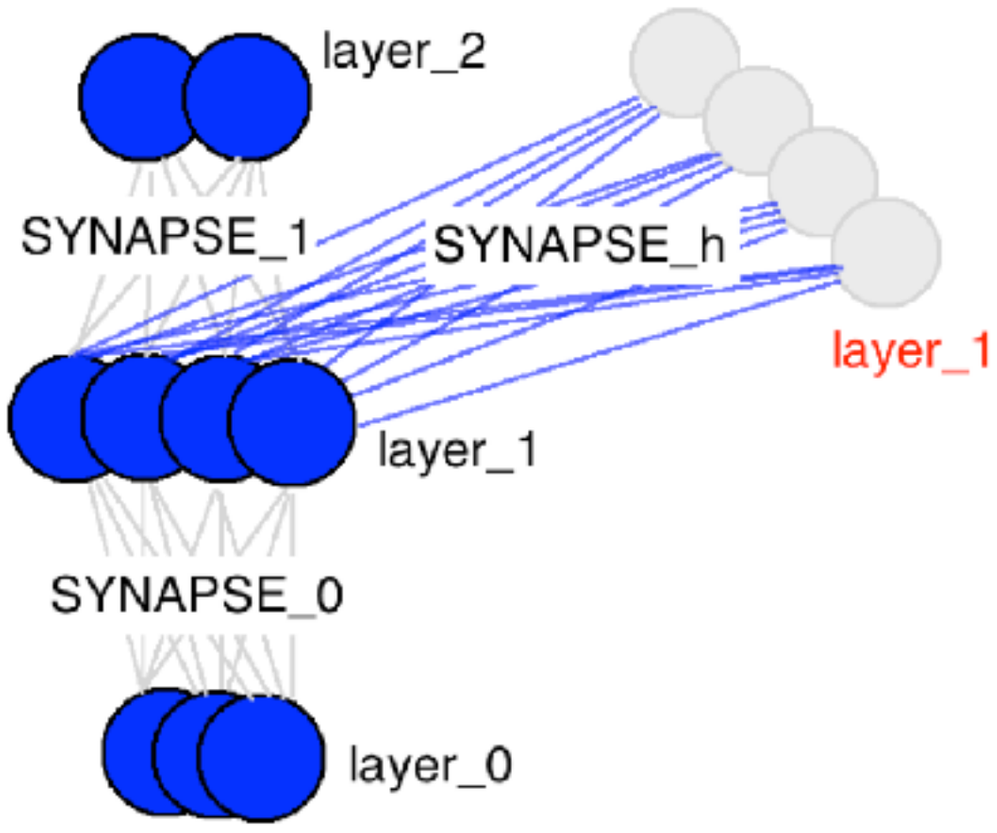


A: neural network
x: input
h: output

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

Recurrent Neural Networks

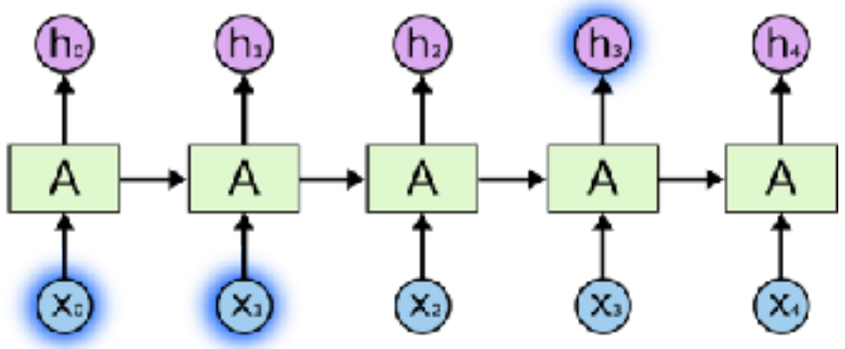
Difference



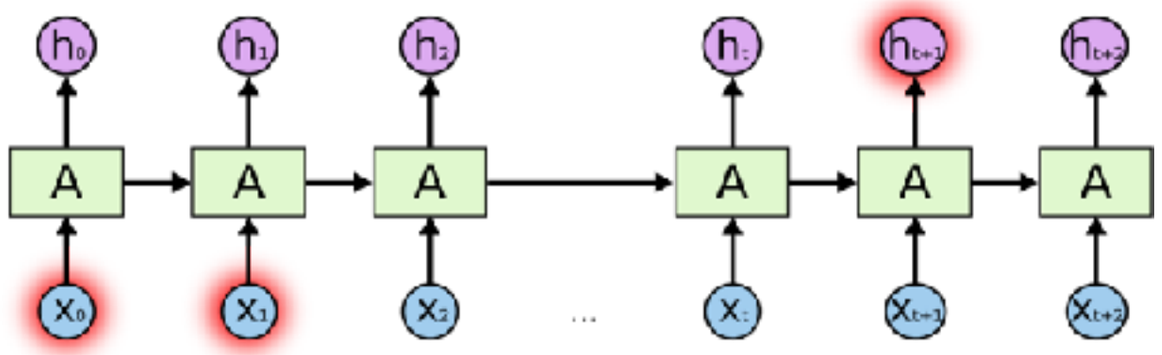
Recurrent Neural Networks

Short- versus long-term memory

The clouds are in the *sky*



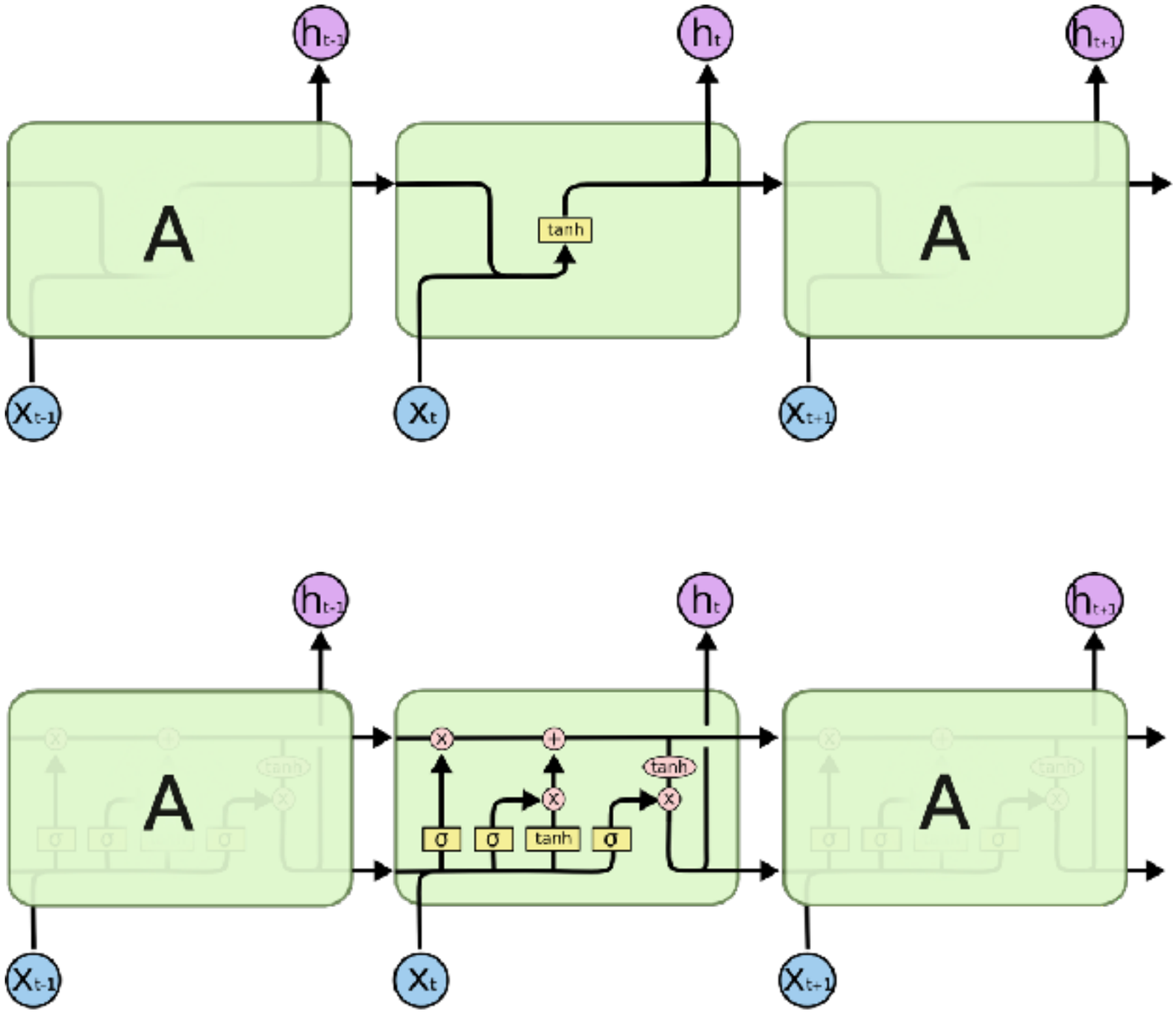
I grew up in France … I speak fluent French.



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

Recurrent Neural Networks

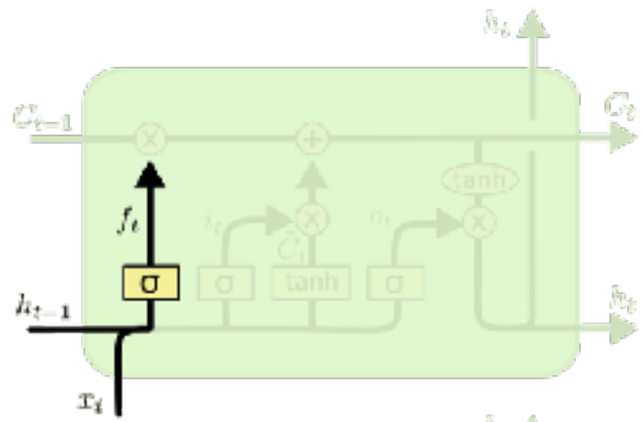
RNN vs LSTM



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

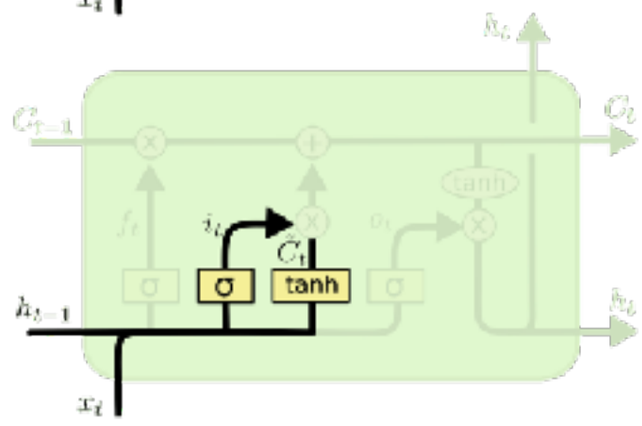
Recurrent Neural Networks

Decide what to forget



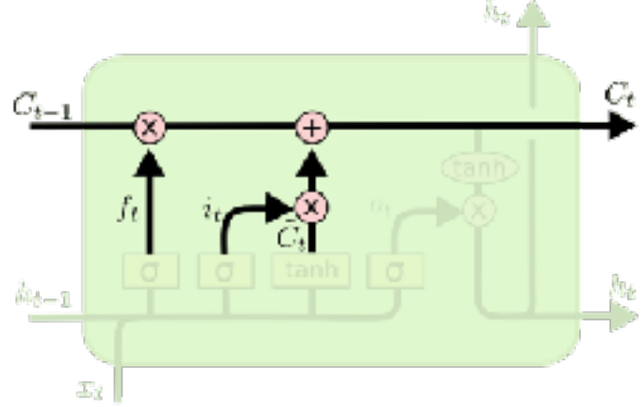
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Decide what to add



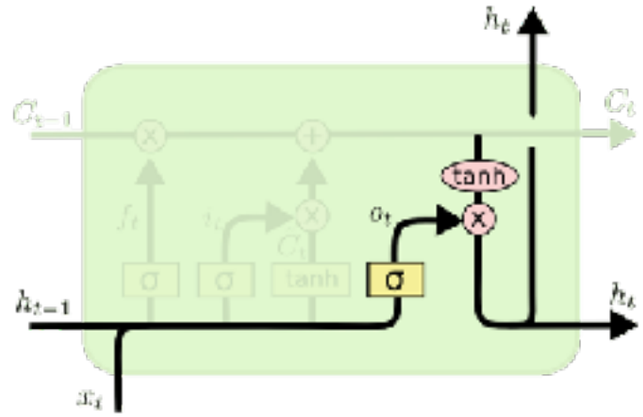
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Update cell state



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

Decide what to output

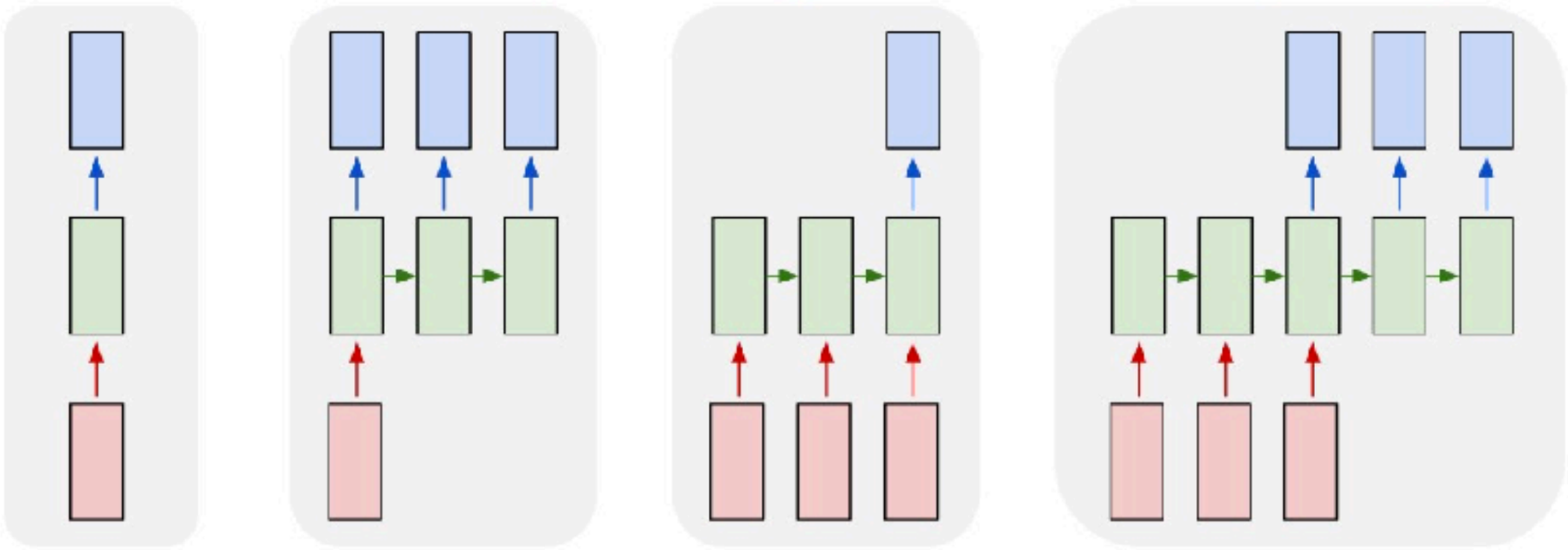


$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

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

Recurrent Neural Networks

Types



sentiment analysis

no RNN

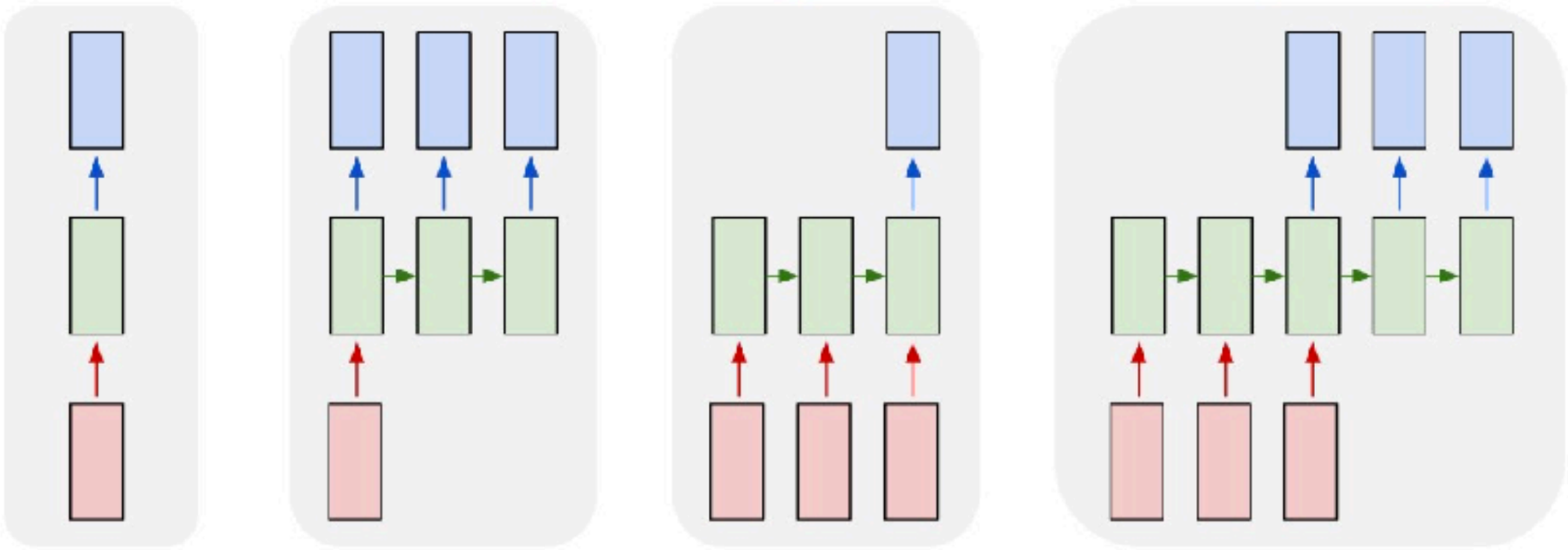
image captioning

language translation

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Recurrent Neural Networks

Types



no RNN image captioning sentiment analysis language translation

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Word Embeddings and Recurrent Neural Networks

Exercise: Text classification

sentiment analysis

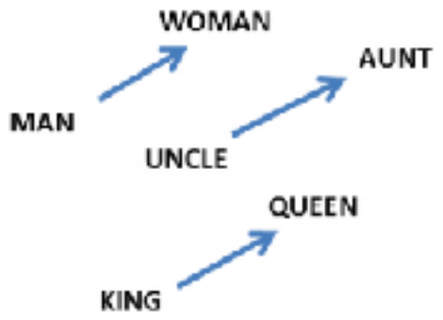
[1]

disease classification

[1] Maas et al., 2011

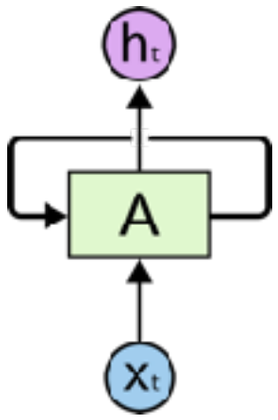
Word Embeddings and Recurrent Neural Networks

A **Word Embedding** is a representation of words where each word is expressed as a real-valued vector. Words with similar meanings have similar vectors.



[2]

A **Recurrent Neural Network** (RNN) is a class of artificial neural network where connections between units form a directed cycle. This allows it to exhibit dynamic temporal behavior. [1]



[3]

[1] https://en.wikipedia.org/wiki/Recurrent_neural_network
[2] Mikolov et al, 2013
[3] http://colah.github.io/posts/2015-08-Understanding-LSTMs