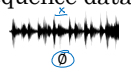
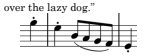
deeplearning.ai

Recurrent Neural Networks

Why sequence models?

1

Examples of sequence data

Speech recognition		"The quick brown fox jumped over the lazy dog."
Music generation	∅	
Sentiment classification	"There is nothing to like in this movie."	★☆☆☆☆
DNA sequence analysis	AGCCCCTGTGAGGAACTAG	AGCCCCTGTGAGGAACTAG
Machine translation	Voulez-vous chanter avec moi?	Do you want to sing with me?
Video activity recognition		Running
Name entity recognition	Yesterday, Harry Potter met Hermione Granger.	Yesterday, Harry Potter met Hermione Granger.

Andrew Ng

2

deeplearning.ai

Recurrent Neural Networks

Notation

3

Motivating example

NLP

x: Harry Potter and Hermione Granger invented a new spell.

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ ..... $x^{<t>}$ ..... $x^{<9>}$

$T_x = 9$

y: 1 1 0 1 1 0 0 0 0

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$ ..... $y^{<9>}$

$T_y = 9$

$x^{(i)<t>}$ $T_x^{(i)} = 9$ 15

$y^{(i)<t>}$ $T_y^{(i)}$

Andrew Ng

4

Representing words

$x^{<t>}$ (x, y)

$x \to y$

x: Harry Potter and Hermione Granger invented a new spell.

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ ... $x^{<t>}$ $x^{<9>}$

Vocabulary

a	1
aaron	2
and	367
harry	4075
potter	6830
zulu	10,000

<UNK> 10,000

$\leftarrow 4075$ $\leftarrow 6830$ $\leftarrow 367$ 10,000

One-hot

Andrew Ng

5

Representing words

x: Harry Potter and Hermione Granger invented a new spell.

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ ... $x^{<9>}$

And = 367
Invented = 4700
A = 1
New = 5976
Spell = 8376
Harry = 4075
Potter = 6830
Hermione = 4200
Gran... = 4000

Andrew Ng

6

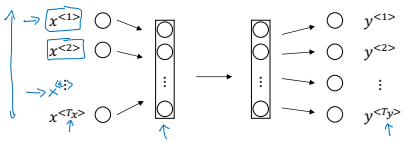
deeplearning.ai

Recurrent Neural Networks

Recurrent Neural Network Model

7

Why not a standard network?



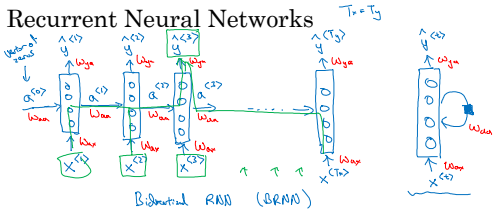
Problems:

- Inputs, outputs can be different lengths in different examples.
- Doesn't share features learned across different positions of text.

Andrew Ng

8

Recurrent Neural Networks



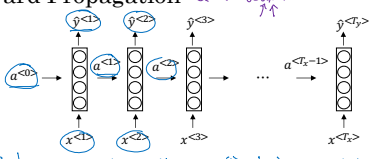
He said, "Teddy Roosevelt was a great President."

He said, "Teddy bears are on sale!"

Andrew Ng

9

Forward Propagation



$a^{<0>} = \vec{0}$

$a^{<1>} = g_1(W_{aa}a^{<0>} + W_{ax}x^{<1>} + b_a)$ ← tanh / ReLU

$\hat{y}^{<1>} = g_2(W_{ya}a^{<1>} + b_y)$ ← sigmoid

$$\boxed{\begin{aligned} a^{<t>} &= g(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a) \\ \hat{y}^{<t>} &= g(W_{ya}a^{<t>} + b_y) \end{aligned}}$$

Andrew Ng

10

Simplified RNN notation

$$a^{<t>} = g(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$

$$\hat{y}^{<t>} = g(W_{ya}a^{<t>} + b_y)$$

$$\hat{y}^{<t>} = g(W_y a^{<t>} + b_y)$$

$$a^{<t>} = g(W_a[a^{<t-1>}, x^{<t>}] + b_a)$$

$$[W_{aa} \vdots W_{ax}] = W_a$$

$$[a^{<t-1>}, x^{<t>}] = \begin{bmatrix} a^{<t-1>} \\ x^{<t>} \end{bmatrix}$$

$$[W_{aa} \vdots W_{ax}] \begin{bmatrix} a^{<t-1>} \\ x^{<t>} \end{bmatrix} = W_{aa}a^{<t-1>} + W_{ax}x^{<t>}$$

Andrew Ng

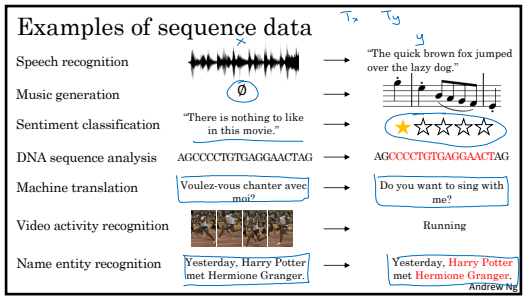
11

deeplearning.ai

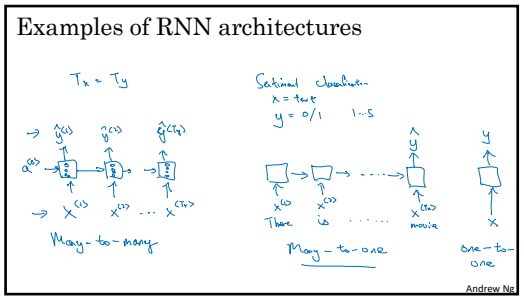
Recurrent Neural Networks

Different types of RNNs

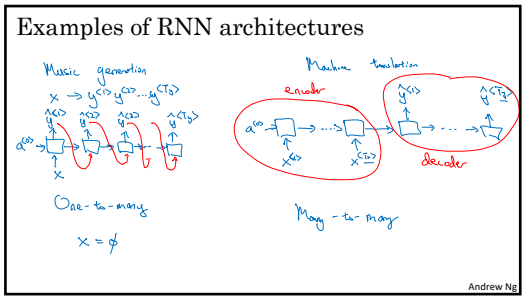
12



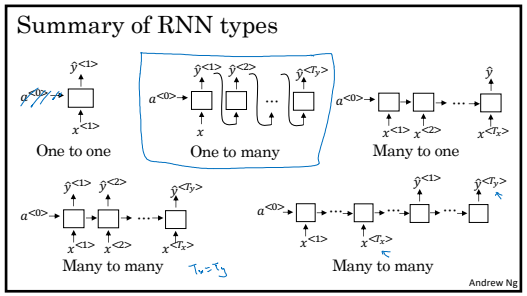
13



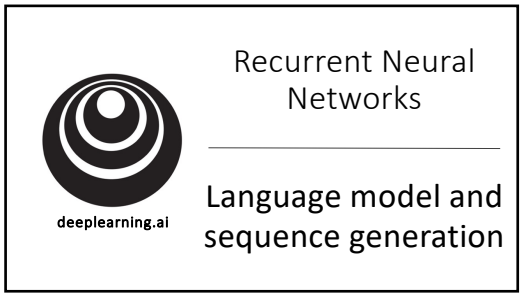
14



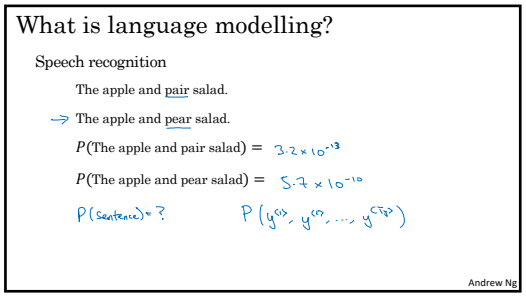
15



16



17



18

Language modelling with an RNN

Training set: large corpus of english text.

Tokenize

Cats average 15 hours of sleep a day. <EOS>

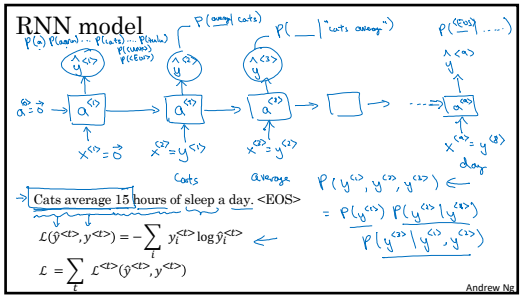
The Egyptian Mau is a bread of cat. <EOS>

10,000 <UNK>

Andrew Ng

19

RNN model



20

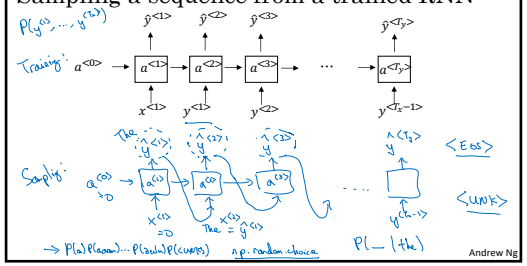
deeplearning.ai

Recurrent Neural Networks

Sampling novel sequences

21

Sampling a sequence from a trained RNN

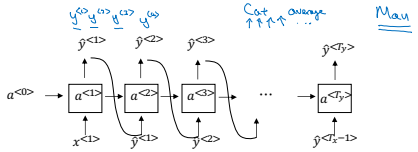


22

Character-level language model

Vocabulary = [a, aaron, …, zulu, <UNK>]

Vocabulary = [a, b, c, …, z, ␣, ., ,, ;, 0, …, 9, A, …, Z]



23

Sequence generation

News

President enrique peña nieto, announced sench's sulk former coming football langston paring.

"I was not at all surprised," said hich langston.

"Concussion epidemic", to be examined.

The gray football the told some and this has on the uefa icon, should money as.

Shakespeare

The mortal moon hath her eclipse in love.

And subject of this thou art another this fold.

When besser be my love to me see sabl's.

For whose are ruse of mine eyes heaves.

Andrew Ng

24

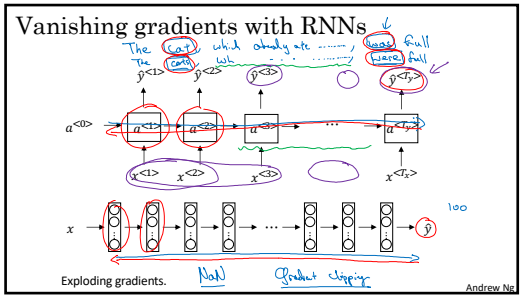
deeplearning.ai

Recurrent Neural Networks

Vanishing gradients with RNNs

25

Vanishing gradients with RNNs



26

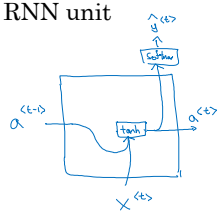
deeplearning.ai

Recurrent Neural Networks

Gated Recurrent Unit (GRU)

27

RNN unit



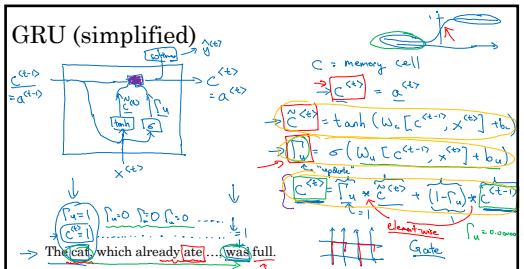
$$a^{<t>} = g(W_a[a^{<t-1>}, x^{<t>}] + b_a)$$

tanh

Andrew Ng

28

GRU (simplified)



29

Full GRU

$$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) + c^{<t-1>}$$

LSTM

The cat, which ate already, was full.

Andrew Ng

30

deeplearning.ai

Recurrent Neural Networks

LSTM (long short term memory) unit

31

GRU and LSTM

GRU

$$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

$$a^{<t>} = c^{<t>}$$

Γ_f

LSTM

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$

(update) $\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u)$

(forget) $\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f)$

(output) $\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o)$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * c^{<t>}$$

Hochreiter & Schmidhuber 1997. Long short-term memory

Andrew Ng

32

LSTM in pictures

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u)$$

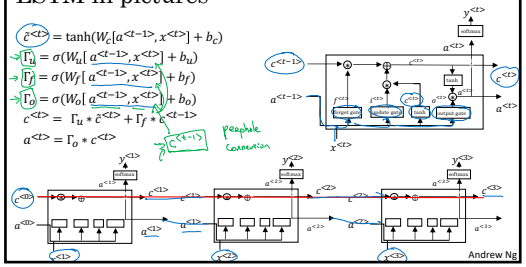
$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f)$$

$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * c^{<t>}$$

Peephole connection



Andrew Ng

34

deeplearning.ai

Recurrent Neural Networks

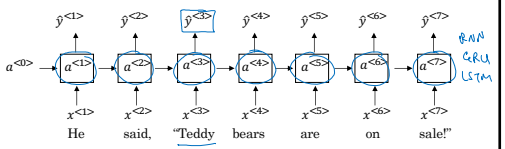
Bidirectional RNN

35

Getting information from the future

He said, "Teddy bears are on sale!"

He said, "Teddy Roosevelt was a great President!"

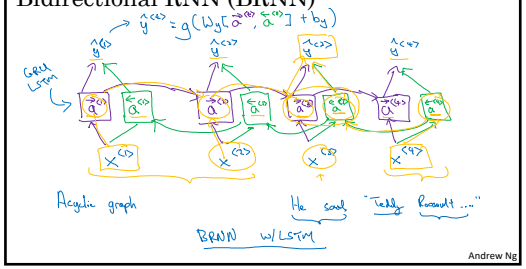


Andrew Ng

36

Bidirectional RNN (BRNN)

$$\hat{y}^{<t>} = g(W_y[\vec{a}^{<t>}, \overleftarrow{a}^{<t>}] + b_y)$$



Andrew Ng

37

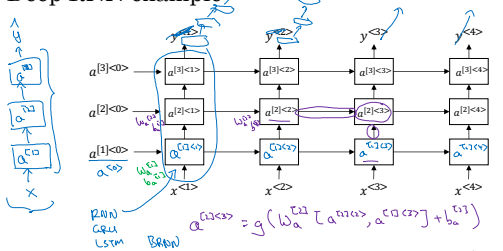
deeplearning.ai

Recurrent Neural Networks

Deep RNNs

38

Deep RNN example



Andrew Ng

39

deeplearning.ai

Sequence to sequence models

Basic models

40

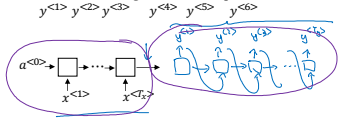
Sequence to sequence model

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ $x^{<4>}$ $x^{<5>}$

Jane visite l'Afrique en septembre

→ Jane is visiting Africa in September.

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$ $y^{<4>}$ $y^{<5>}$ $y^{<6>}$



[Sutskever et al., 2014. Sequence to sequence learning with neural networks]
[Cho et al., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation]

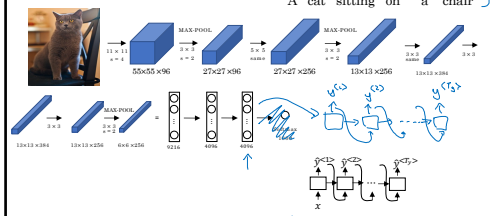
Andrew Ng

41

Image captioning

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$ $y^{<4>}$ $y^{<5>}$ $y^{<6>}$

A cat sitting on a chair



Mao et. al., 2014. Deep captioning with multimodal recurrent neural networks]
Vinyals et. al., 2014. Show and tell: Neural image caption generator]
Karpathy and Li, 2015. Deep visual-semantic alignments for generating image descriptions]

Andrew Ng

42

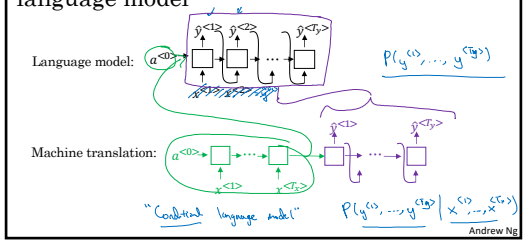
deeplearning.ai

Sequence to sequence models

Picking the most likely sentence

43

Machine translation as building a conditional language model



44

Finding the most likely translation

Jane visite l'Afrique en septembre.

$P(y^{<1>}, \ldots, y^{<T_y>} \mid x)$

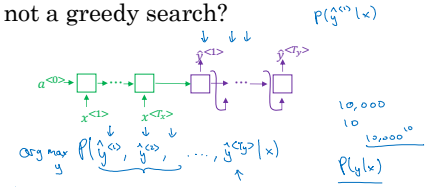
- Jane is visiting Africa in September.
- Jane is going to be visiting Africa in September.
- In September, Jane will visit Africa.
- Her African friend welcomed Jane in September.

$$\underset{y^{<1>},\ldots,y^{<T_y>}}{\arg\max}\ P(y^{<1>}, \ldots, y^{<T_y>} \mid x)$$

Andrew Ng

45

Why not a greedy search?



- Jane is visiting Africa in September.
- Jane is going to be visiting Africa in September.

Andrew Ng

46

deeplearning.ai

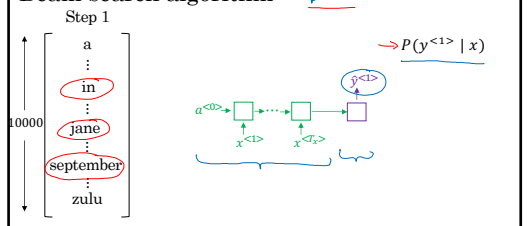
Sequence to sequence models

Beam search

47

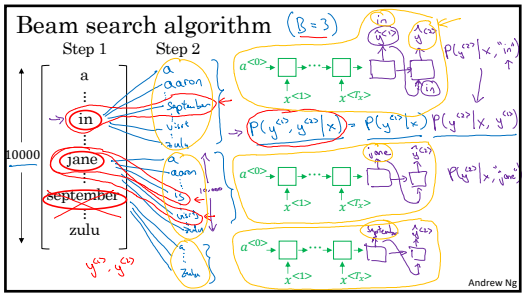
Beam search algorithm

B = 3 (beam width)

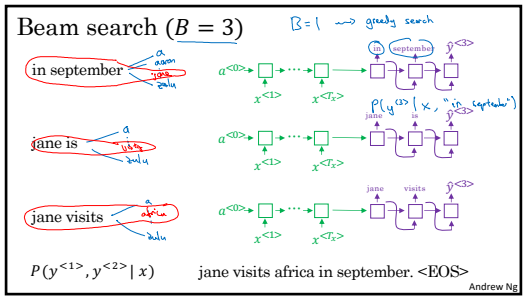


48

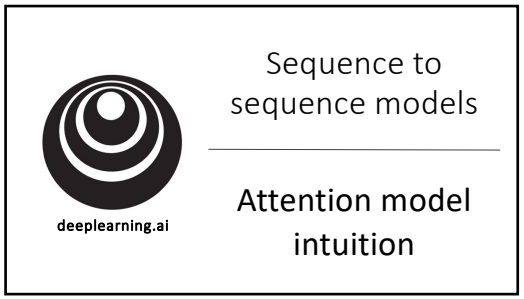
Beam search algorithm (B = 3)



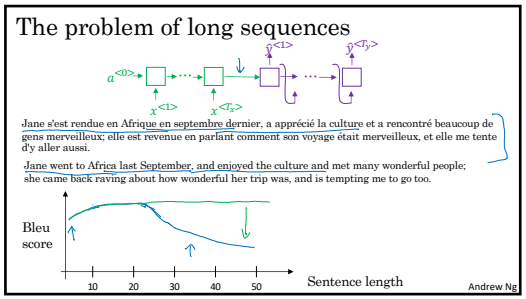
49



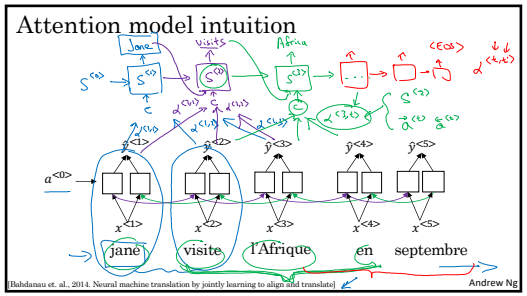
50



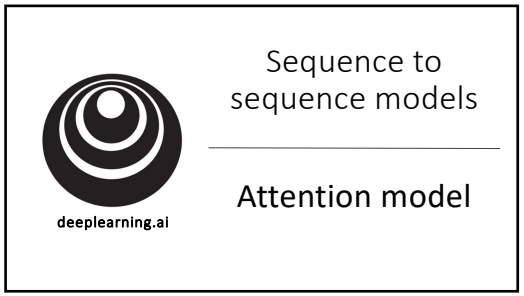
51



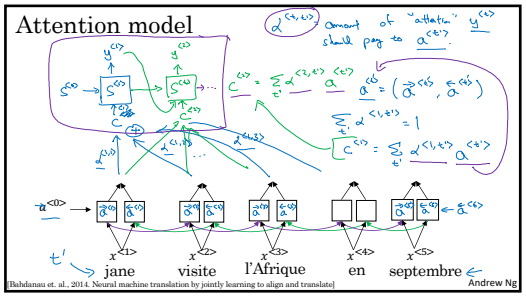
52



53



54



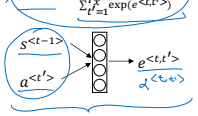
55

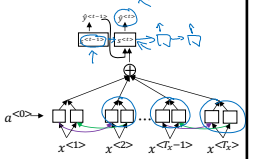
Computing attention $\alpha^{<t,t'>}$

T_x T_y

$\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

$$\alpha^{<t,t'>} = \frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$$





[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]
[Xu et. al., 2015. Show, attend and tell: Neural image caption generation with visual attention]

Andrew Ng

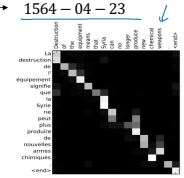
56

Attention examples

July 20th 1969 → 1969 – 07 – 20

23 April, 1564 → 1564 – 04 – 23

Visualization of $\alpha^{<t,t'>}$:



Andrew Ng

57