Sequence to sequence models

Basic models

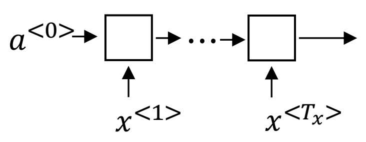
deeplearning.ai

1

Sequence to sequence model

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ $x^{<4>}$ $x^{<5>}$
Jane visite l'Afrique en septembre

$\longrightarrow$ Jane is visiting Africa in September.
$y^{<1>}$ $y^{<2>}$ $y^{<3>}$ $y^{<4>}$ $y^{<5>}$ $y^{<6>}$

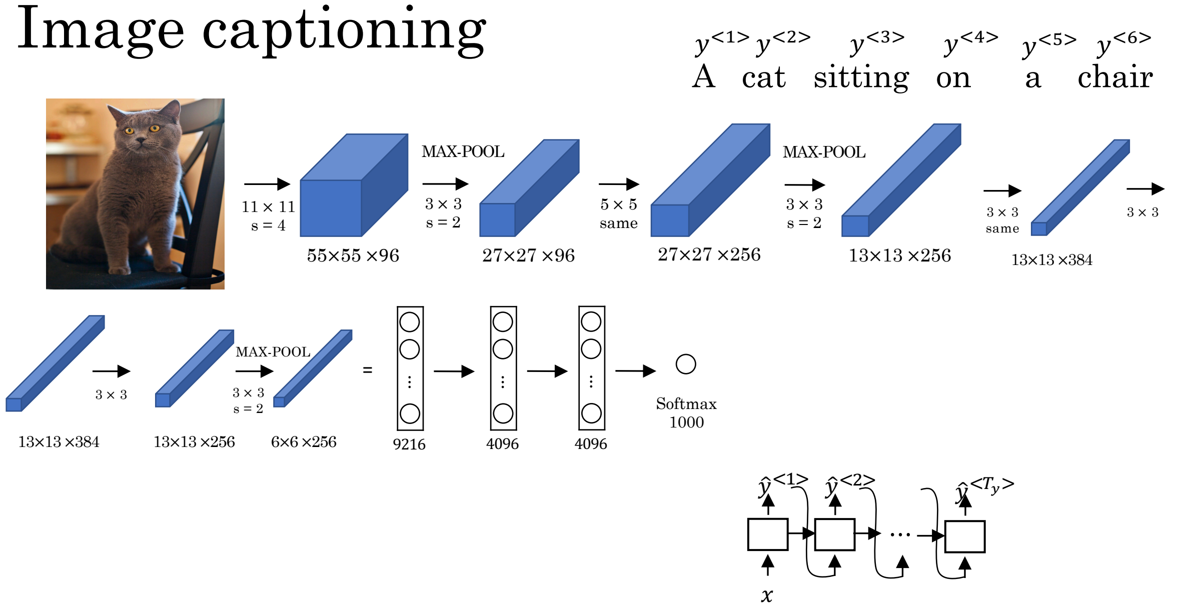


[Sutskever et al., 2014. Sequence to sequence learning with neural networks]
[Cho et al., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation]

Andrew Ng

2

Image captioning



[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

3

deeplearning.ai

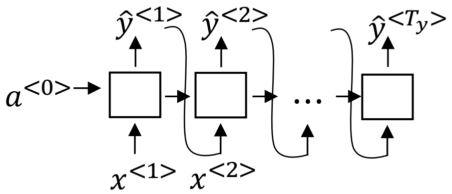
Sequence to sequence models

Picking the most likely sentence

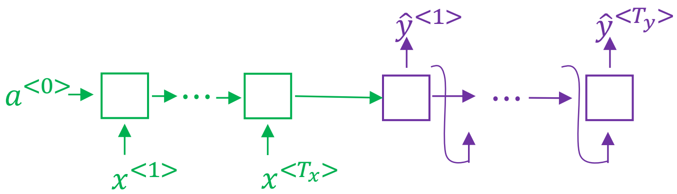
4

Machine translation as building a conditional language model

Language model:



Machine translation:



Andrew Ng

5

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.

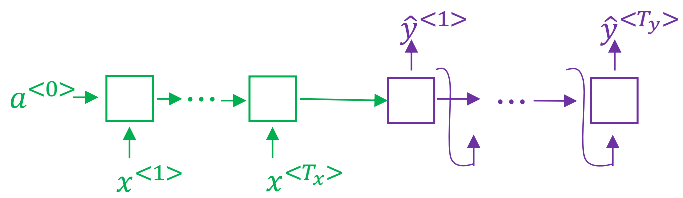
→ Her African friend welcomed Jane in September.

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

Andrew Ng

6

Why not a greedy search?



→ Jane is visiting Africa in September.

→ Jane is going to be visiting Africa in September.

Andrew Ng

7

deeplearning.ai

Sequence to sequence models

Beam search

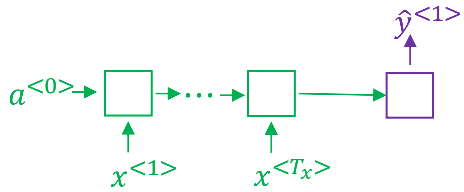
8

Beam search algorithm

Step 1

$$10000 \; \begin{bmatrix} \text{a} \\ \vdots \\ \text{in} \\ \vdots \\ \text{jane} \\ \vdots \\ \text{september} \\ \vdots \\ \text{zulu} \end{bmatrix}$$

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



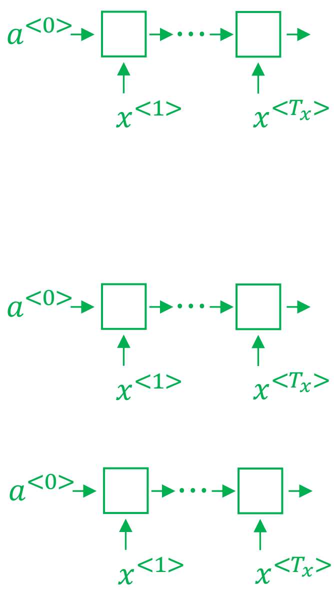
Andrew Ng

9

Beam search algorithm

Step 1 Step 2

$$10000 \; \begin{bmatrix} \text{a} \\ \vdots \\ \text{in} \\ \vdots \\ \text{jane} \\ \vdots \\ \text{september} \\ \vdots \\ \text{zulu} \end{bmatrix}$$



Andrew Ng

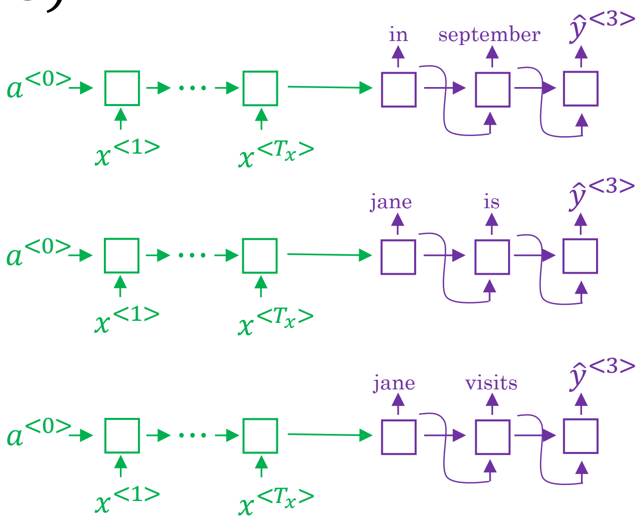
10

Beam search ($B = 3$)

in september

jane is

jane visits



$P(y^{<1>}, y^{<2>} | x)$

jane visits africa in september. <EOS>

Andrew Ng

11

deeplearning.ai

Sequence to sequence models

Refinements to beam search

12

Length normalization

$$\arg\max_{y} \prod_{t=1}^{T_y} P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>})$$

$$\arg\max_{y} \sum_{t=1}^{T_y} \log P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>})$$

$$\sum_{t=1}^{T_y} \log P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>})$$

Andrew Ng

13

Beam search discussion

Beam width B?

Unlike exact search algorithms like BFS (Breadth First Search) or DFS (Depth First Search), Beam Search runs faster but is not guaranteed to find exact maximum for $\arg\max_{y} P(y|x)$.

Andrew Ng

14

Sequence to sequence models

deeplearning.ai

Error analysis on beam search

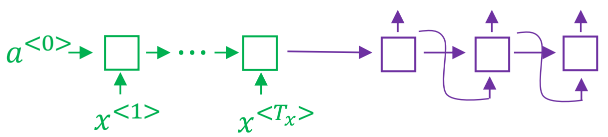
15

Example

Jane visite l'Afrique en septembre.

Human: Jane visits Africa in September.

Algorithm: Jane visited Africa last September.



Andrew Ng

16

Error analysis on beam search

Human: Jane visits Africa in September. (y^*)

Algorithm: Jane visited Africa last September. ($\hat{y}$)

Case 1:

Beam search chose $\hat{y}$. But y^* attains higher $P(y|x)$.

Conclusion: Beam search is at fault.

Case 2:

y^* is a better translation than $\hat{y}$. But RNN predicted $P(y^*|x) < P(\hat{y}|x)$.

Conclusion: RNN model is at fault.

Andrew Ng

17

Error analysis process

Human	Algorithm	$P(y^*\|x)$	$P(\hat{y}\|x)$	At fault?
Jane visits Africa in September.	Jane visited Africa last September.			

Figures out what faction of errors are "due to" beam search vs. RNN model

Andrew Ng

18

deeplearning.ai

Sequence to sequence models

Bleu score (optional)

19

Evaluating machine translation

French: Le chat est sur le tapis.

Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: the the the the the the the.

Precision: Modified precision:

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Andrew Ng

20

Bleu score on bigrams

Example: Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: The cat the cat on the mat.

the cat

cat the

cat on

on the

the mat

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Andrew Ng

21

Bleu score on unigrams

Example: Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: The cat the cat on the mat.

$$p_1 = \frac{\sum_{unigram \in \hat{y}} count_{clip}\,(unigram)}{\sum_{unigram \in \hat{y}} count\,(unigram)} \qquad p_n = \frac{\sum_{ngram \in \hat{y}} count_{clip}\,(ngram)}{\sum_{ngram \in \hat{y}} count\,(ngram)}$$

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Andrew Ng

22

Bleu details

p_n = Bleu score on n-grams only

Combined Bleu score:

$$\text{BP} = \begin{cases} 1 & \text{if MT\_output\_length} > \text{reference\_output\_length} \\ \exp(1 - \text{MT\_output\_length}/\text{reference\_output\_length}) & \text{otherwise} \end{cases}$$

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Andrew Ng

23

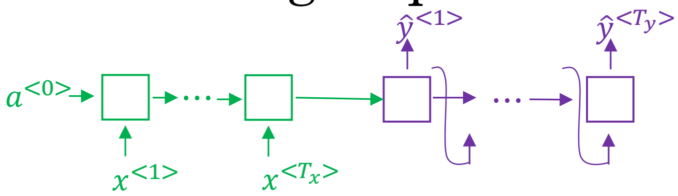
deeplearning.ai

Sequence to sequence models

Attention model intuition

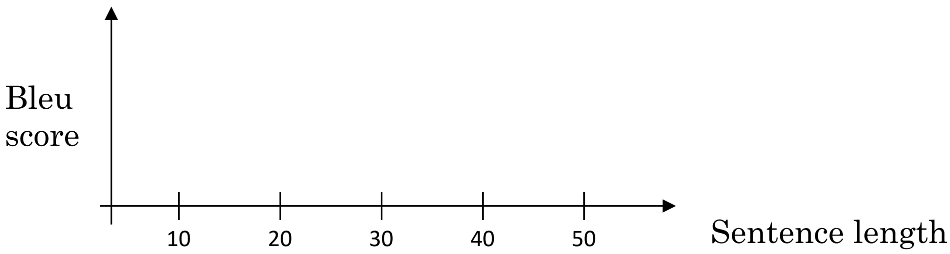
24

The problem of long sequences



Jane s'est rendue en Afrique en septembre dernier, a apprécié la culture et a rencontré beaucoup de gens merveilleux; elle est revenue en parlant comment son voyage était merveilleux, et elle me tente d'y aller aussi.

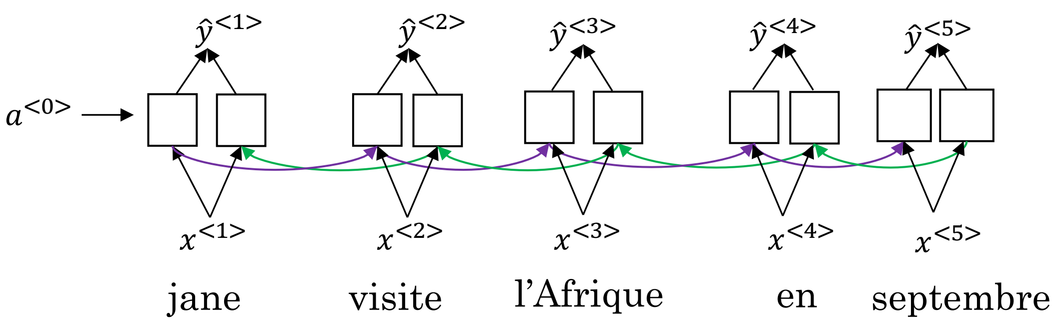
Jane went to Africa last September, and enjoyed the culture and met many wonderful people; she came back raving about how wonderful her trip was, and is tempting me to go too.



Andrew Ng

25

Attention model intuition



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Andrew Ng

26

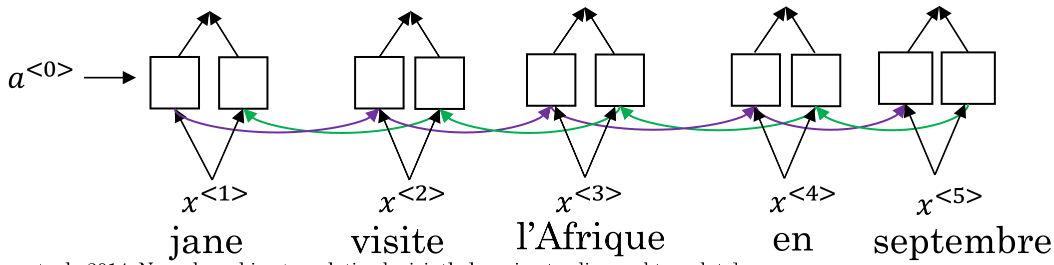
Sequence to sequence models

deeplearning.ai

Attention model

27

Attention model



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

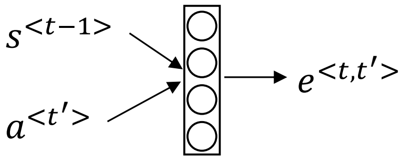
Andrew Ng

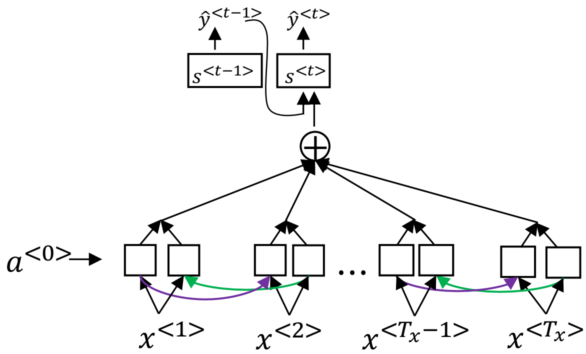
28

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

$\alpha^{<t,t'>}$ = 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

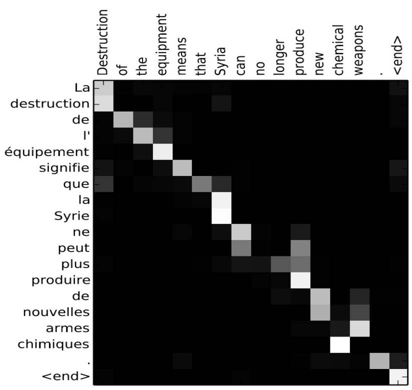
29

Attention examples

July 20th 1969 ⟶ 1969 – 07 – 20

23 April, 1564 ⟶ 1564 – 04 – 23

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



Andrew Ng

30

deeplearning.ai

Audio data

Speech recognition

31

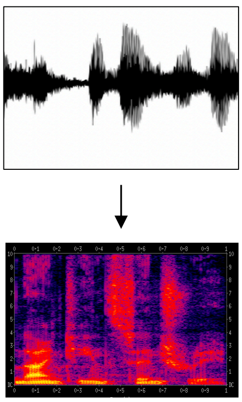
Speech recognition problem

x

audio clip

y

transcript

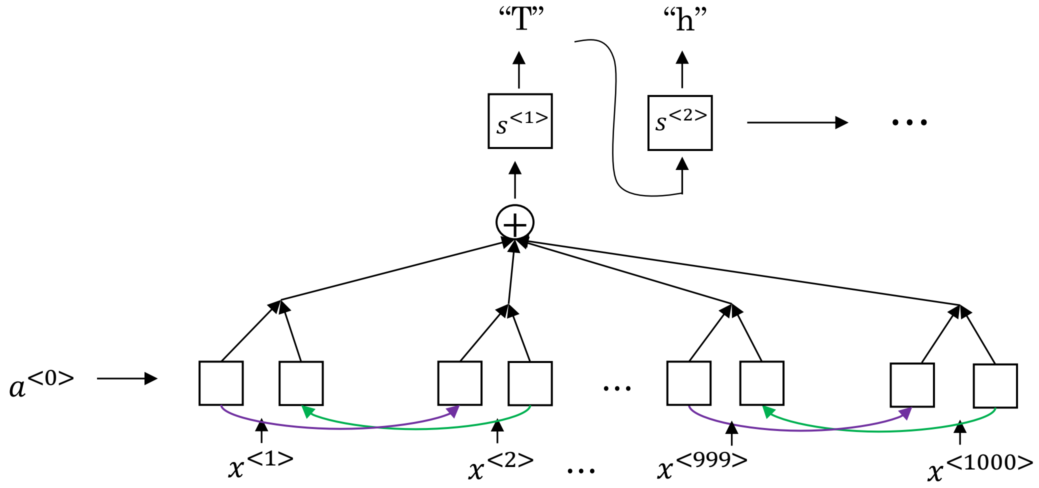


"the quick brown fox"

Andrew Ng

32

Attention model for speech recognition



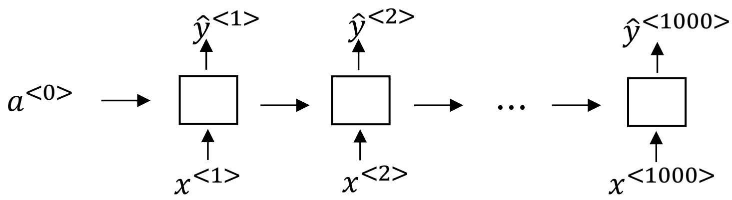
Andrew Ng

33

CTC cost for speech recognition

(Connectionist temporal classification)

"the quick brown fox"



Basic rule: collapse repeated characters not separated by "blank"

[Graves et al., 2006. Connectionist Temporal Classification: Labeling unsegmented sequence data with recurrent neural networks] Andrew Ng

34

deeplearning.ai

Audio data

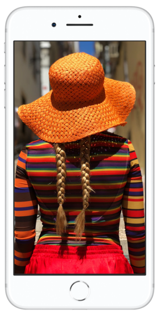
Trigger word detection

35

What is trigger word detection?

Amazon Echo
(Alexa)

Baidu DuerOS
(xiaodunihao)



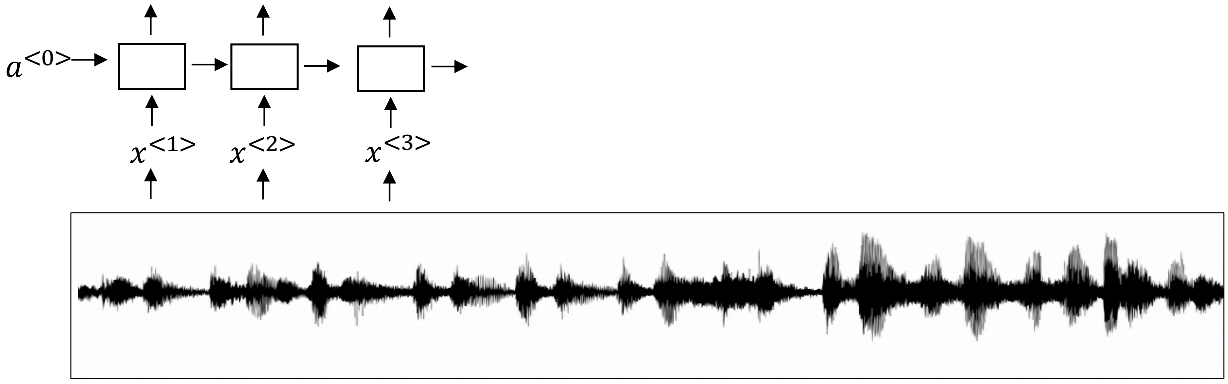
Apple Siri
(Hey Siri)

Google Home
(Okay Google)

Andrew Ng

36

Trigger word detection algorithm



Andrew Ng

37

deeplearning.ai

Conclusion

Summary and thank you

38

Specialization outline

1. Neural Networks and Deep Learning
2. Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
3. Structuring Machine Learning Projects
4. Convolutional Neural Networks
5. Sequence Models

Andrew Ng

39

Deep learning is a super power

Please buy this from shutterstock and replace in final video.

www.shutterstock.com · 331201091

Andrew Ng

40

Thank you.

-Andrew Ng

41