Lecture III:
Deep Learning with Memory

李宏毅
Hung-yi Lee

Outline of Lecture III

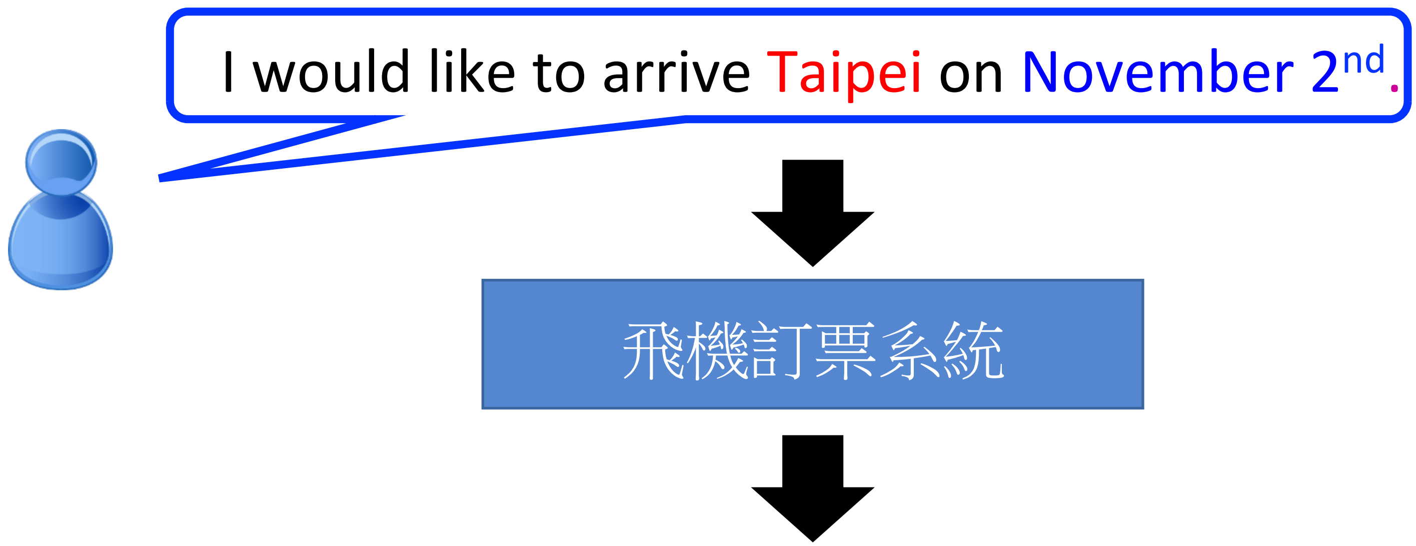
Recurrent Neural Network (RNN) & LSTM

Variants of RNN

Next Wave: Attention–based Model

Example Application

- Slot Filling



Destination: Taipei

time of arrival: November 2nd

Example Application

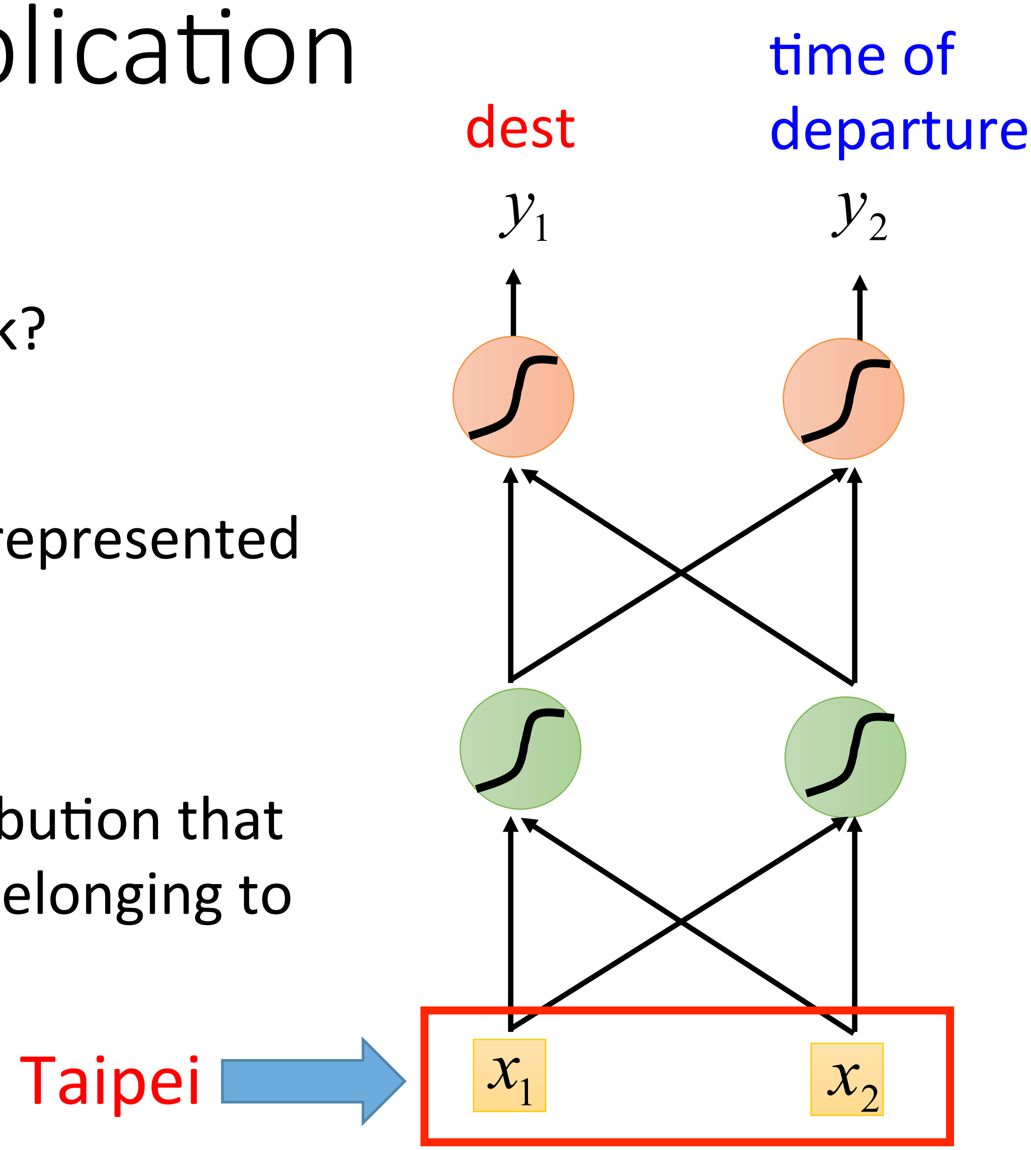
Solving slot filling by Feedforward network?

Input: a word

(Each word is represented as a vector)

Output:

Probability distribution that the input word belonging to the slots



Example Application

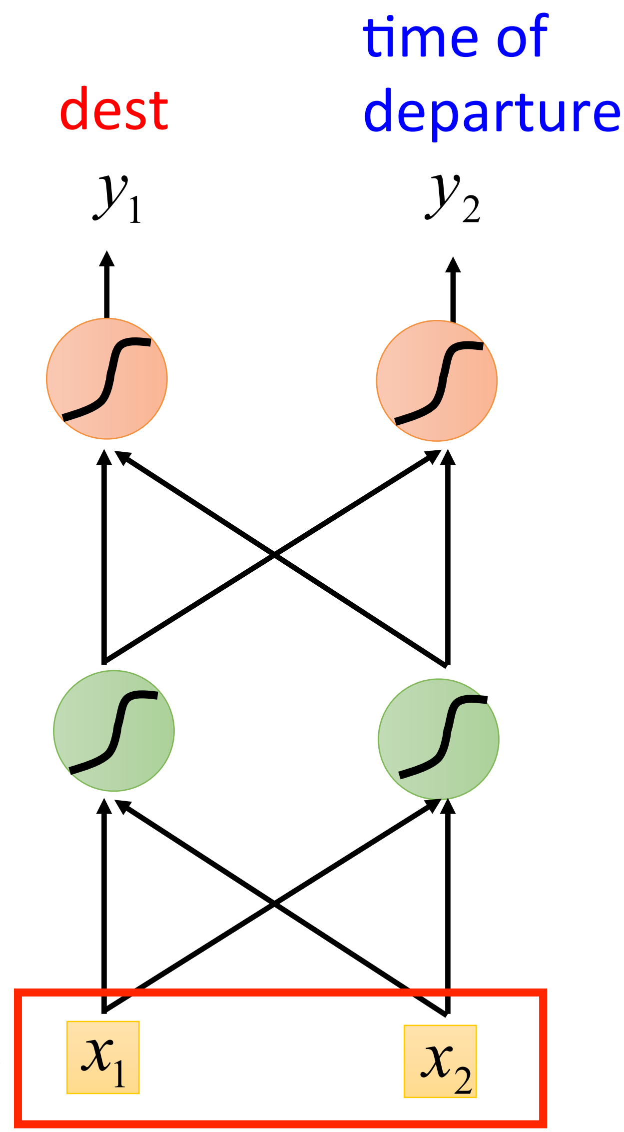
arrive	Taipei	on	November	2^{nd}
other	des	other	time	time

Problem?

leave	Taipei	on	November	2^{nd}
	place of departure		

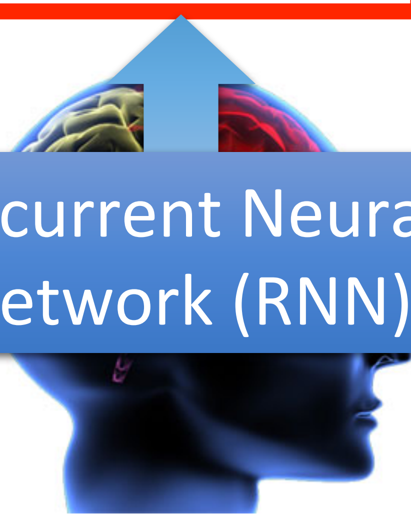
Neural network needs memory!

Taipei



Recurrent Neural Network

Step 1: Neural Network → Step 2: Learning Target → Step 3: Learn!

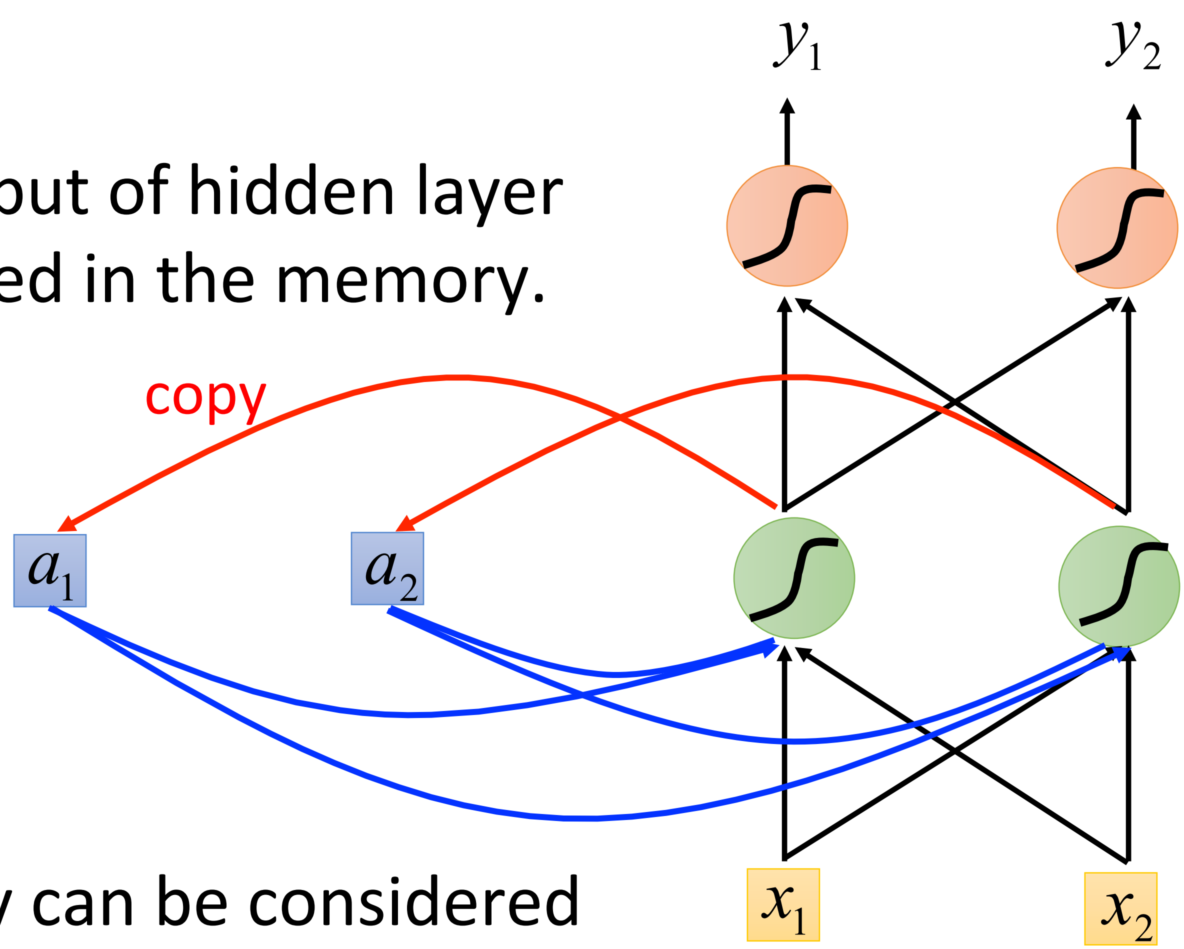


Recurrent Neural Network (RNN)

天生的腦

Recurrent Neural Network (RNN)

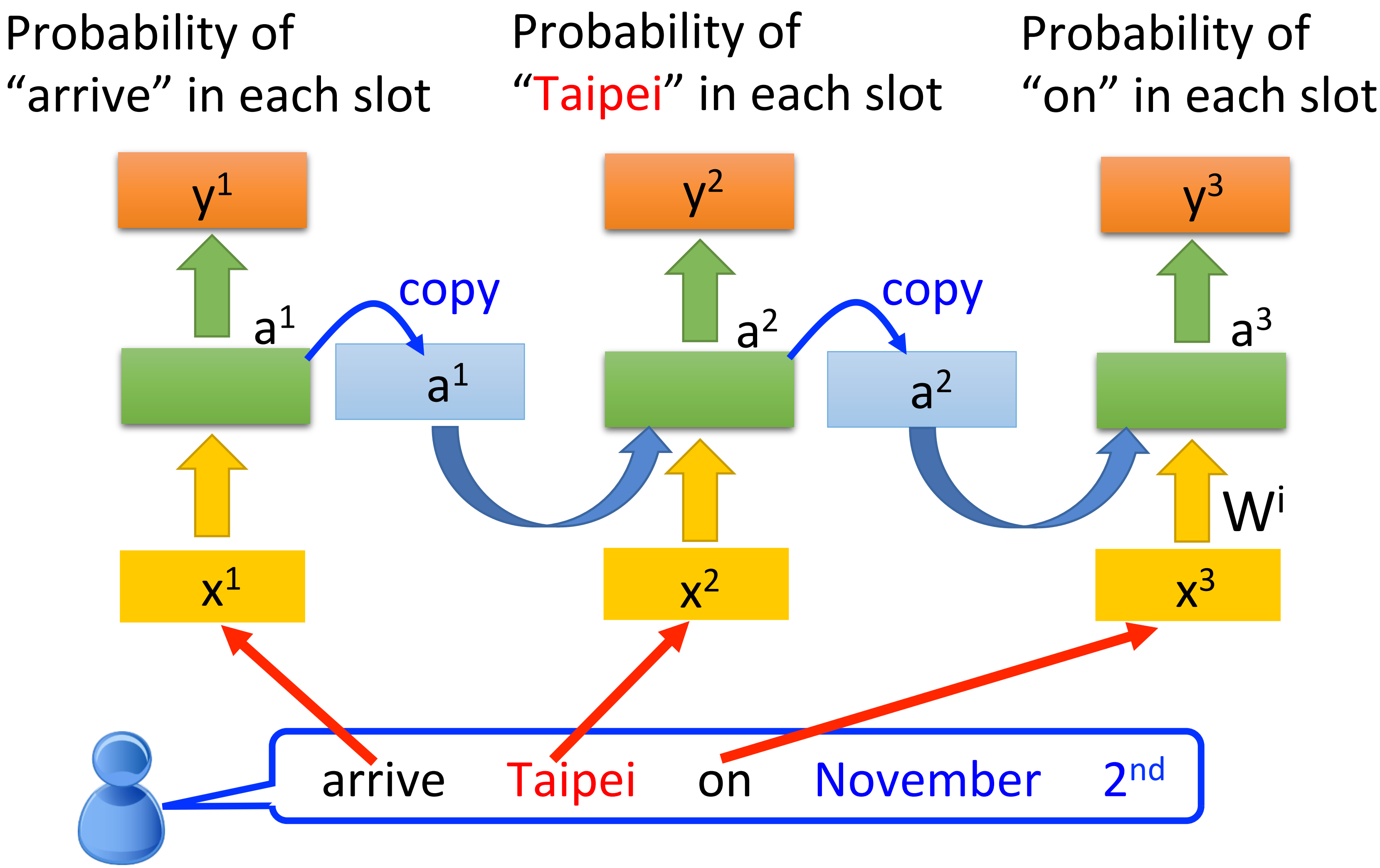
The output of hidden layer are stored in the memory.



Memory can be considered as another input.

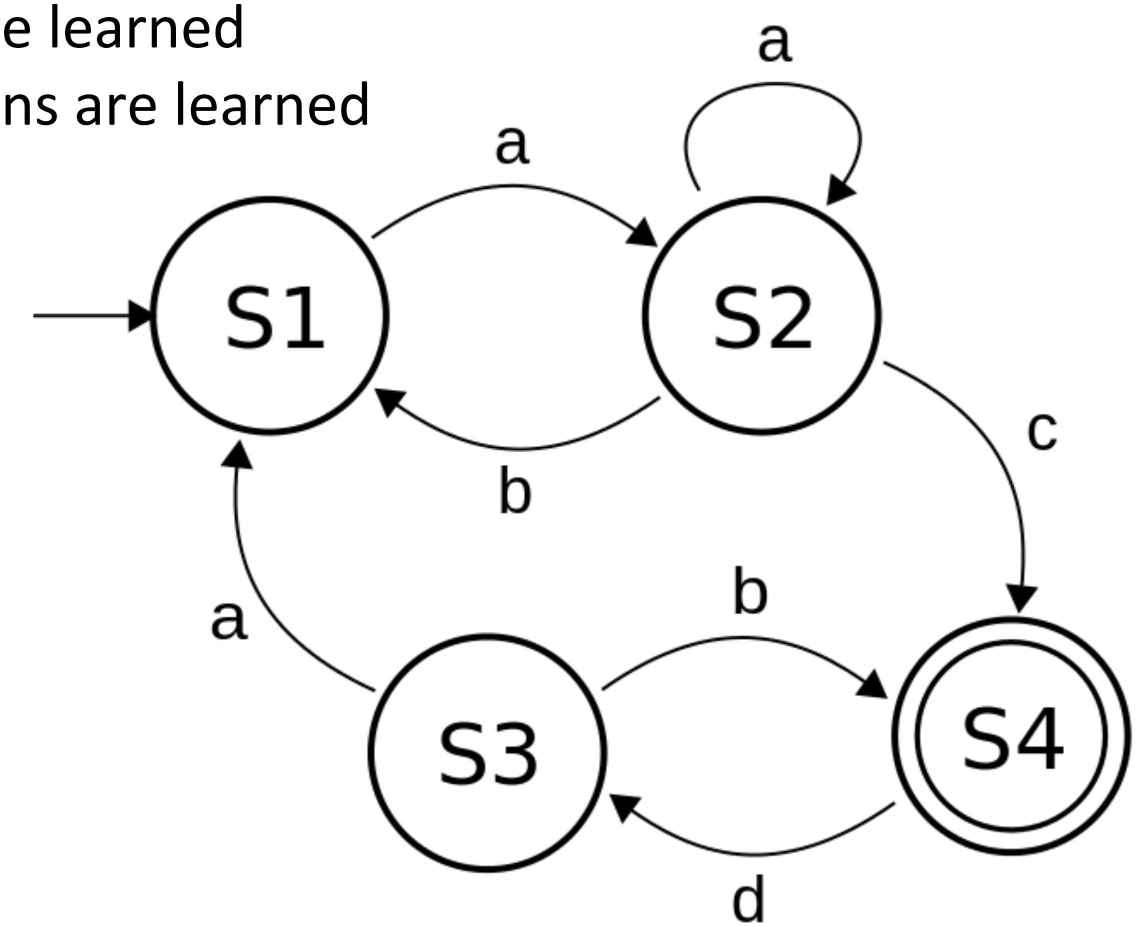
RNN

The same network is used again and again.

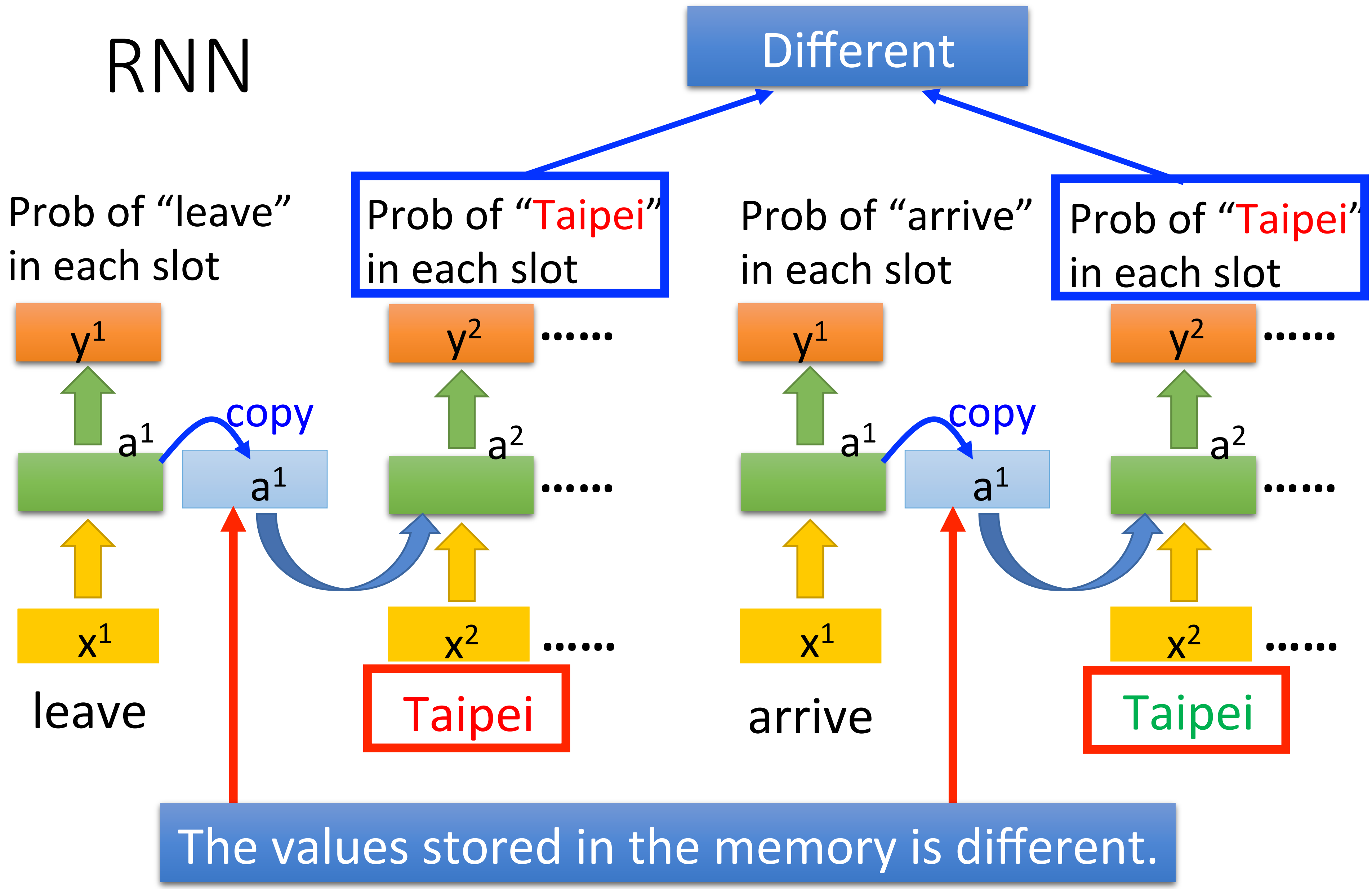


RNN- state machine perspective

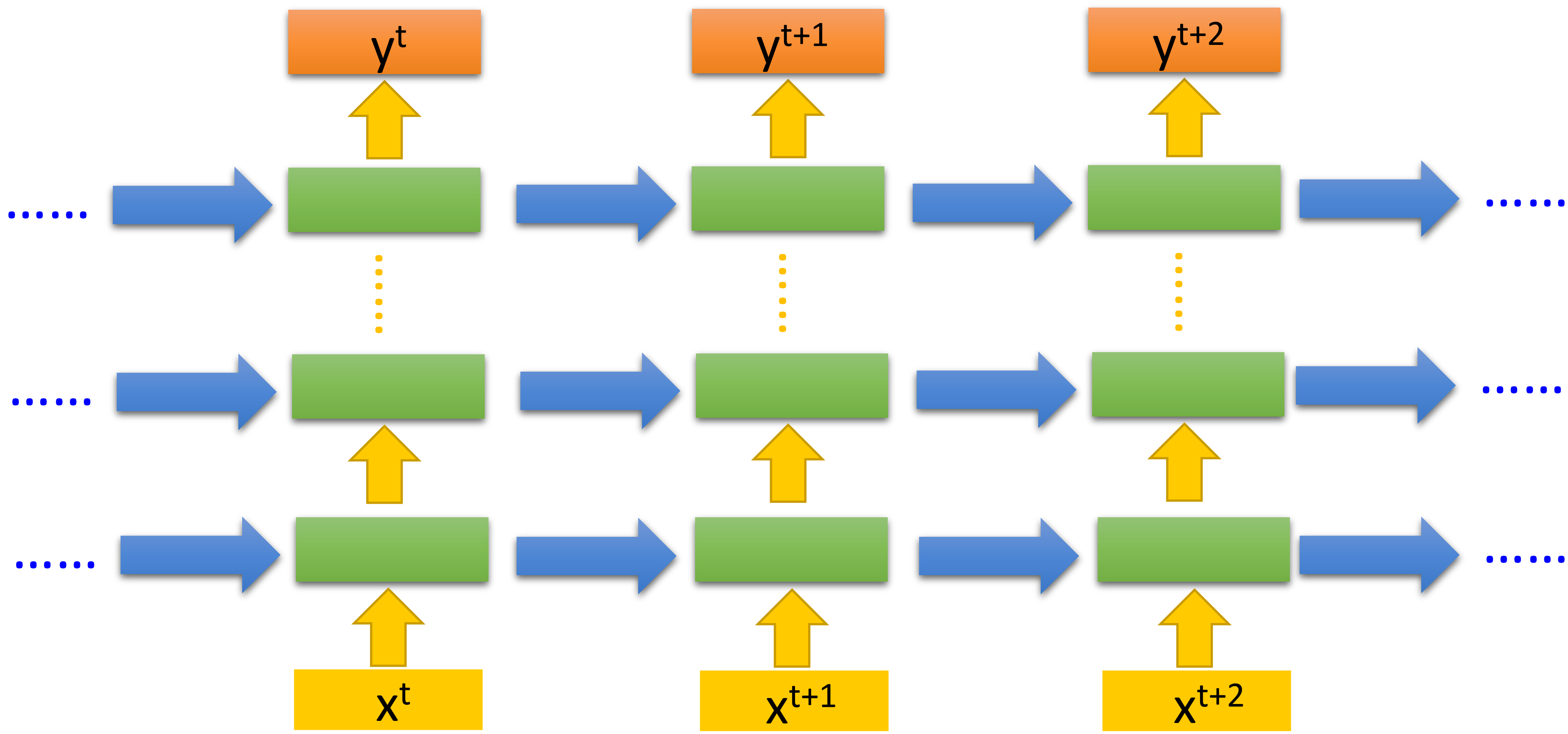
- Infinite number of states
- States are learned
- Transitions are learned



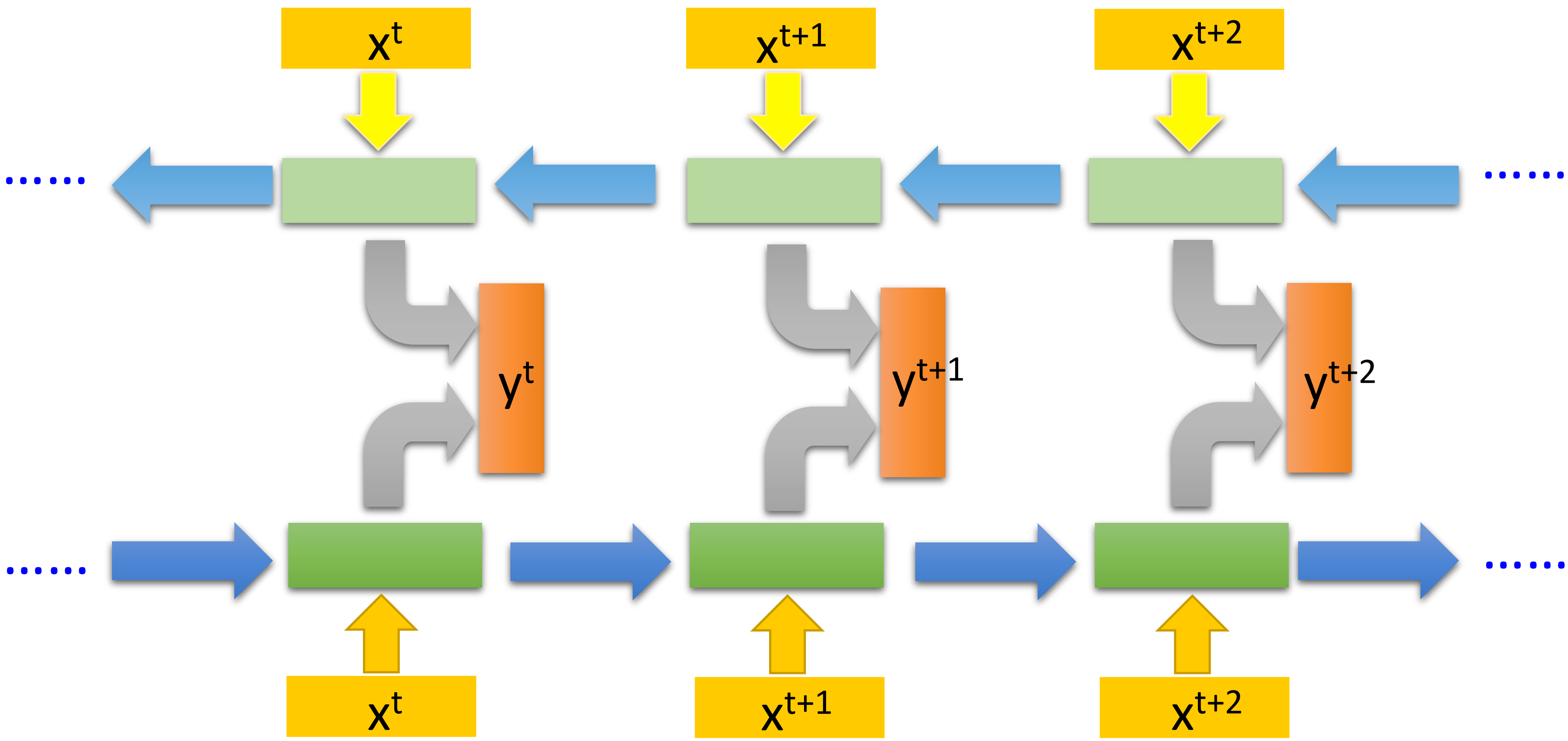
RNN


Different
Prob of "leave" in each slot
Prob of "Taipei" in each slot
Prob of "arrive" in each slot
Prob of "Taipei" in each slot
y^1
y^2 ……
y^1
y^2 ……
a^1
copy
a^1
a^2 ……
a^1
copy
a^1
a^2 ……
x^1
x^2 ……
x^1
x^2 ……
leave
Taipei
arrive
Taipei
The values stored in the memory is different.

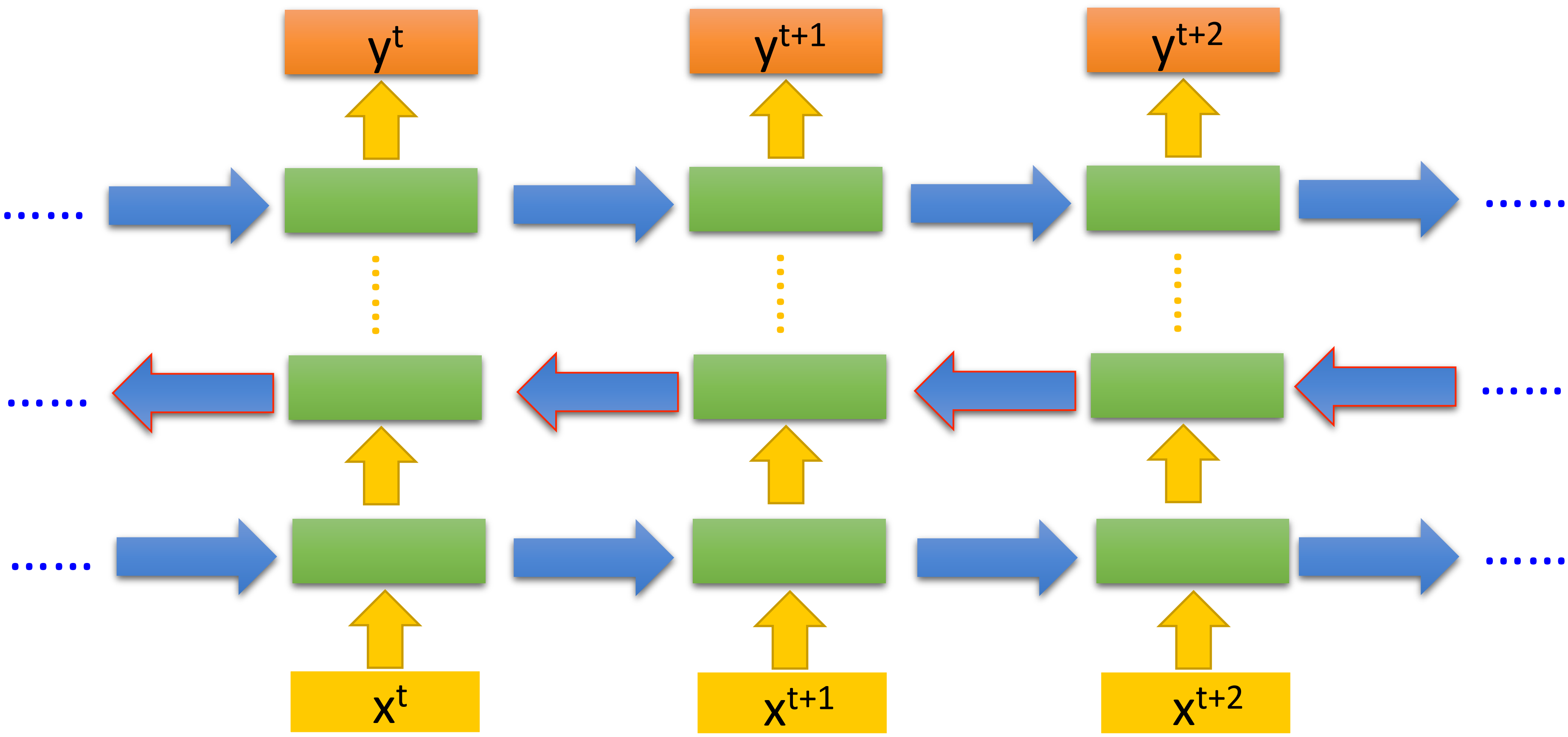
Of course it can be deep …



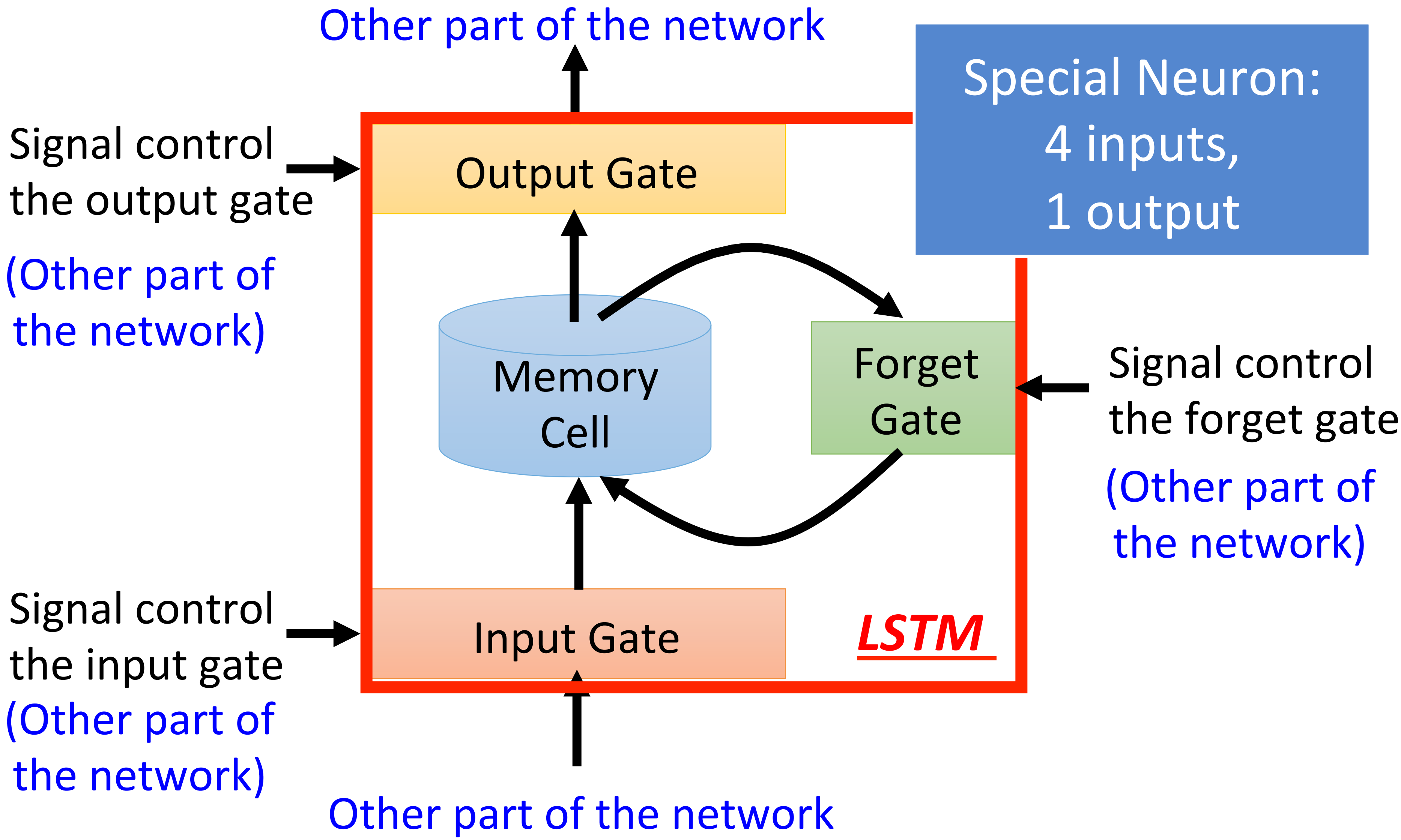
Bidirectional RNN



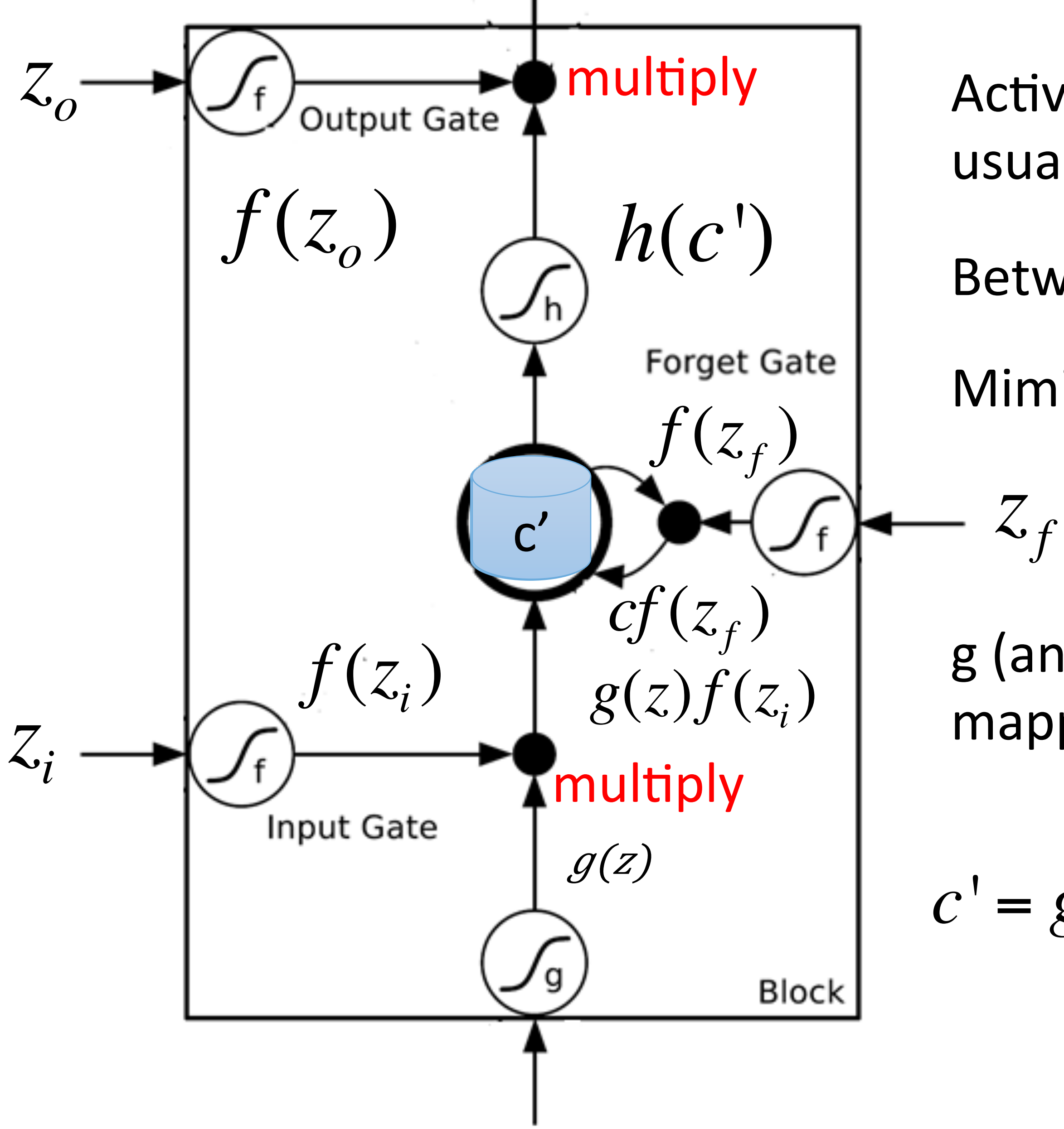
or Deep Bidirectional



Long Short-term Memory (LSTM)



$$a = h(c')f(z_0)$$



Activation function f is usually a sigmoid function

Between 0 and 1

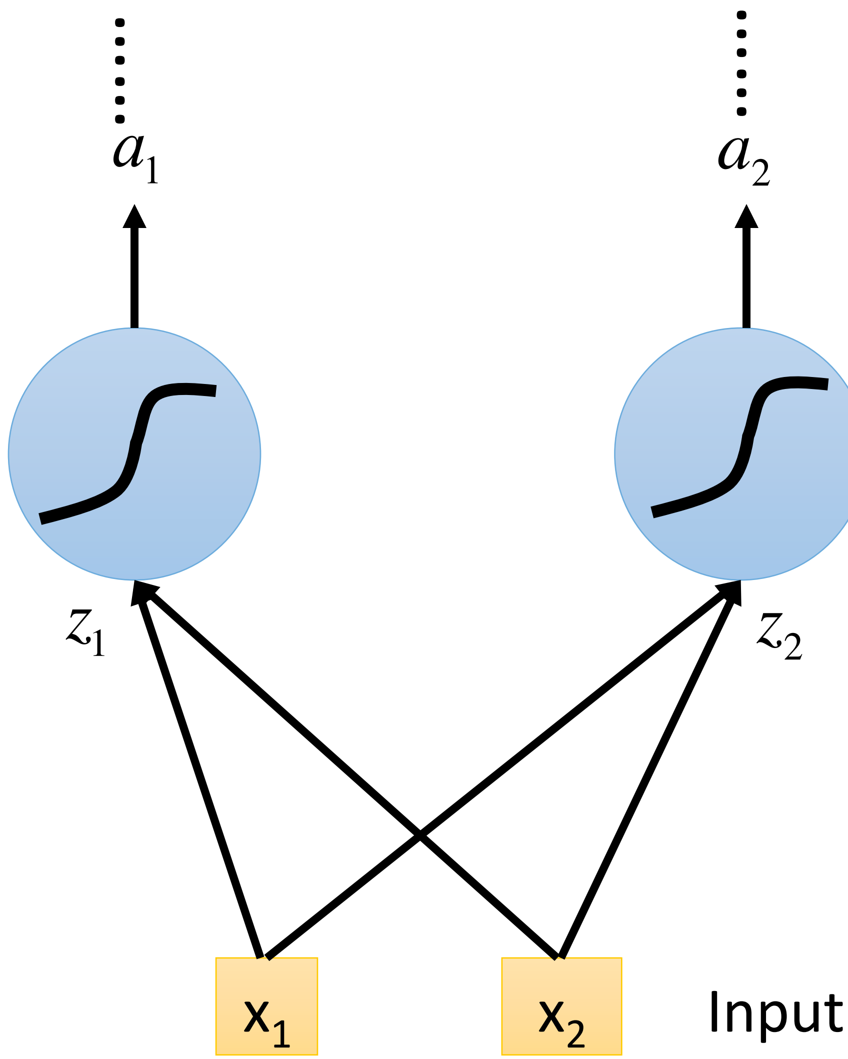
Mimic open and close gate

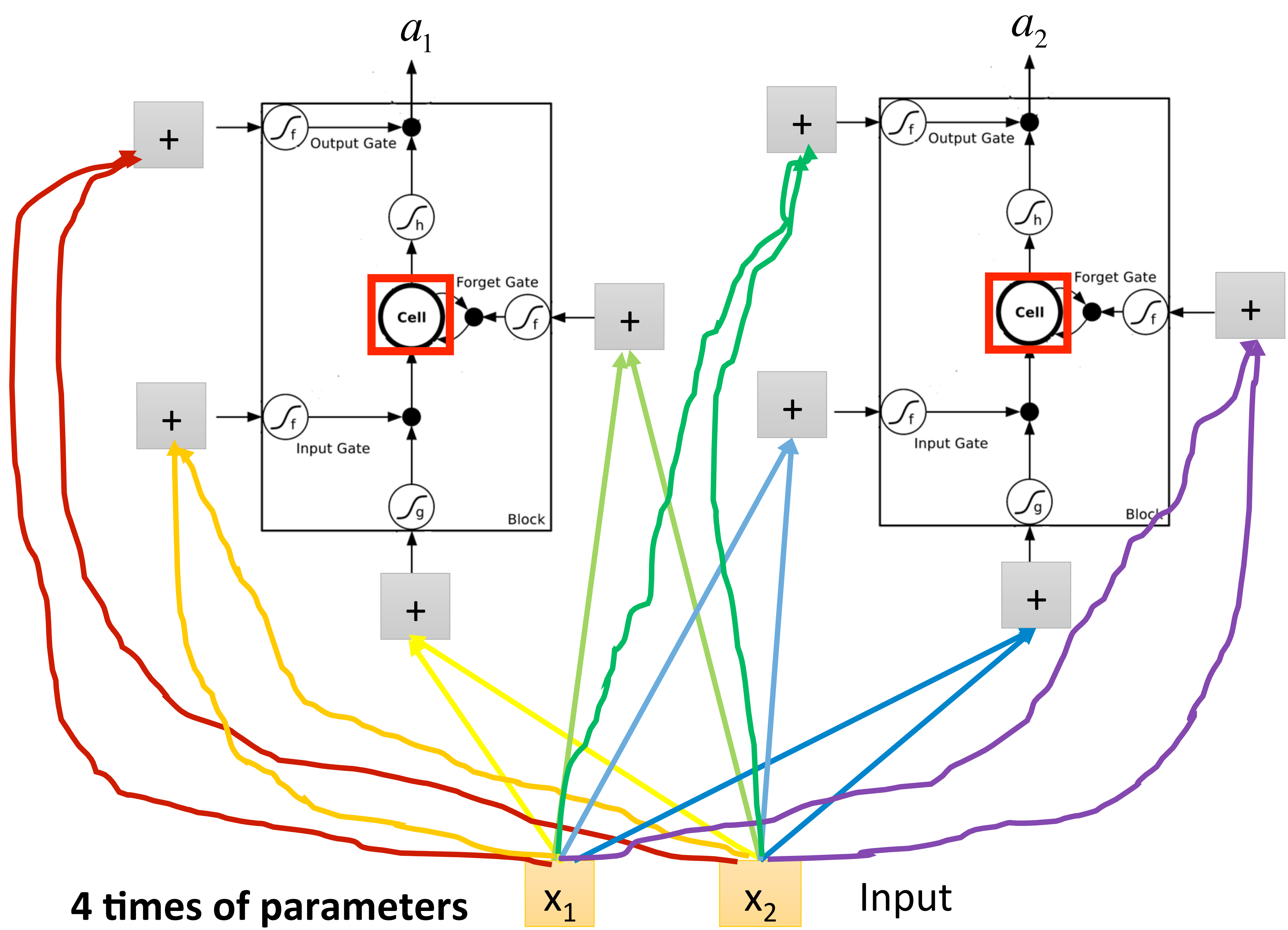
g (and h) is a tanh function mapping between -1 and 1

$$c' = g(z)f(z_i) + cf(z_f)$$

Original Network:

➢Simply replace the neurons with LSTM

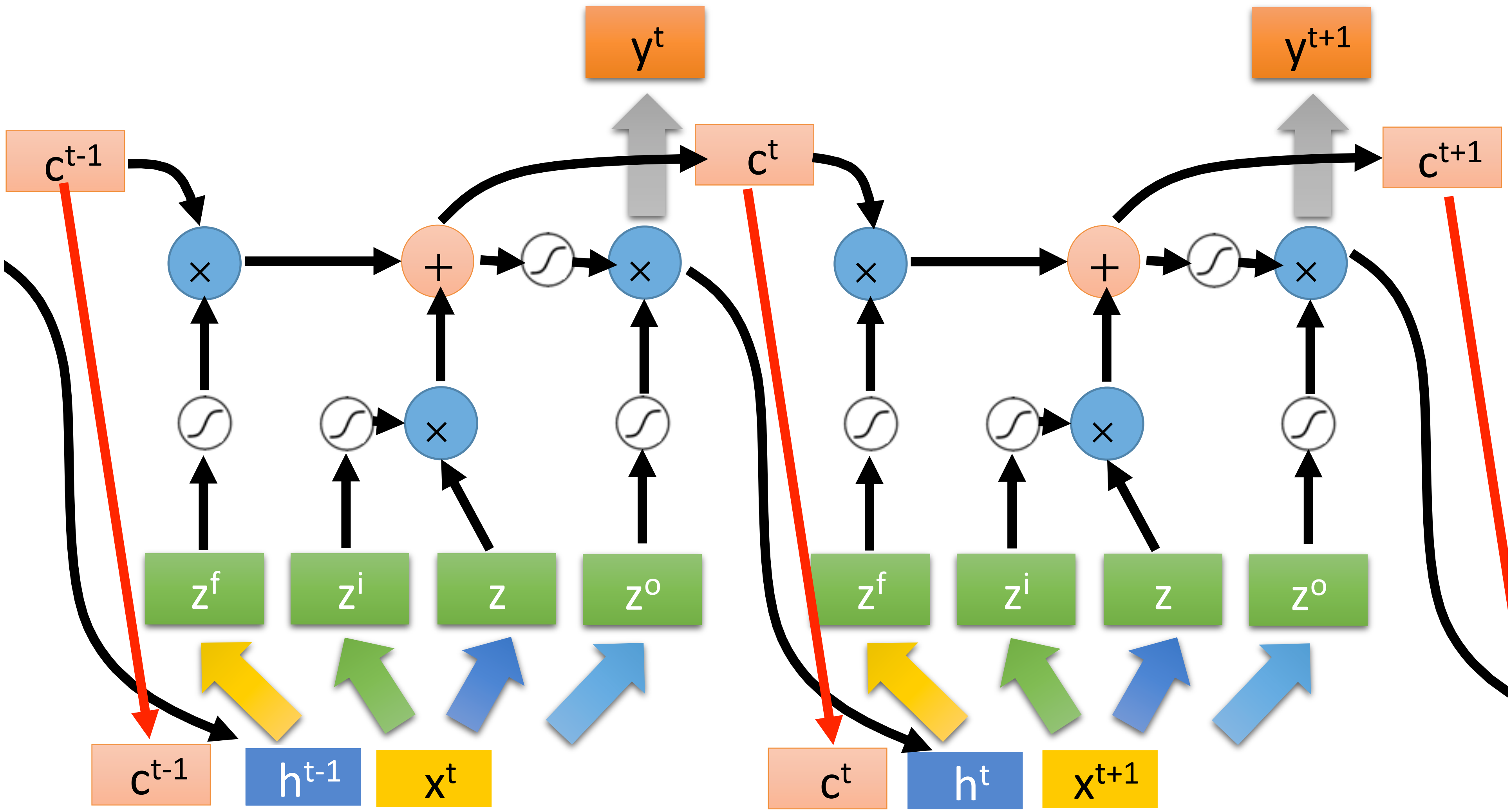




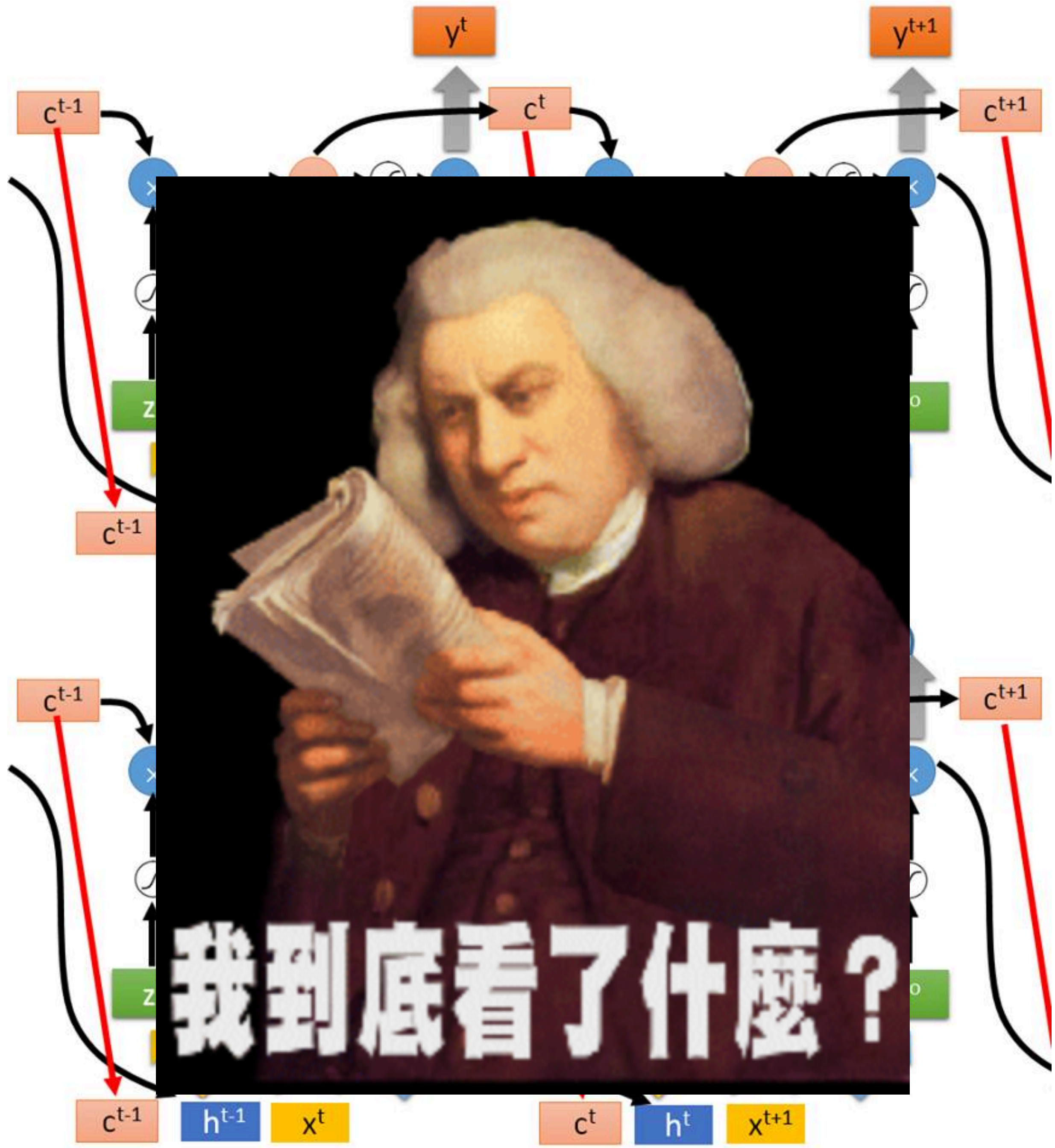
a_1
a_2
+
Output Gate
Forget Gate
Cell
Input Gate
Block
x_1
x_2
Input
4 times of parameters

LSTM

Extension: “peephole”



Multiple-layer LSTM



It is quite standard now.

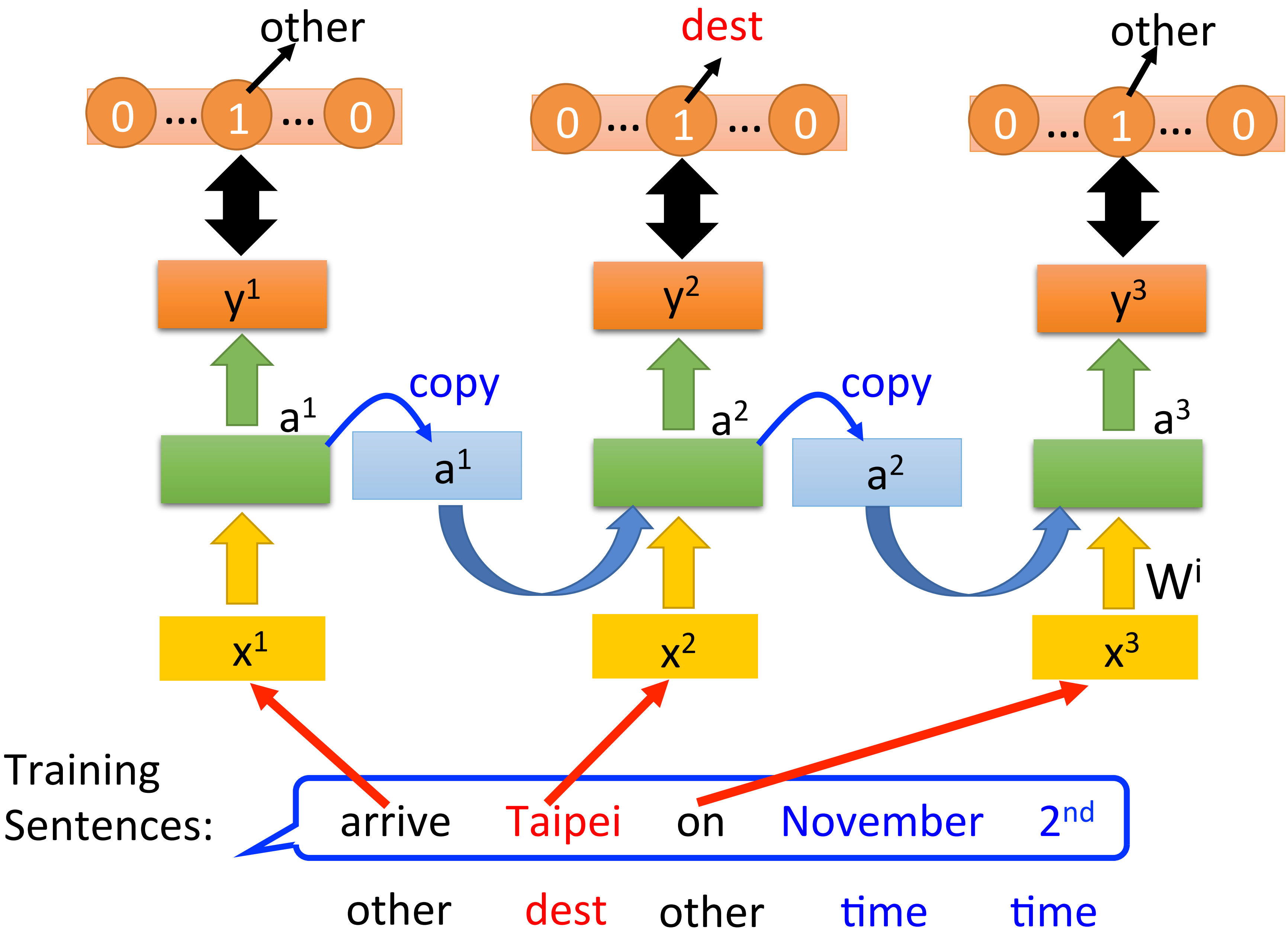
https://img.komicolle.org/2015-09-20/src/14426967627131.gif

Recurrent Neural Network

Step 1: Neural Network → **Step 2: Learning Target** → Step 3: Learn!

天生的腦

Learning Target



Recurrent Neural Network

Step 1: Neural Network

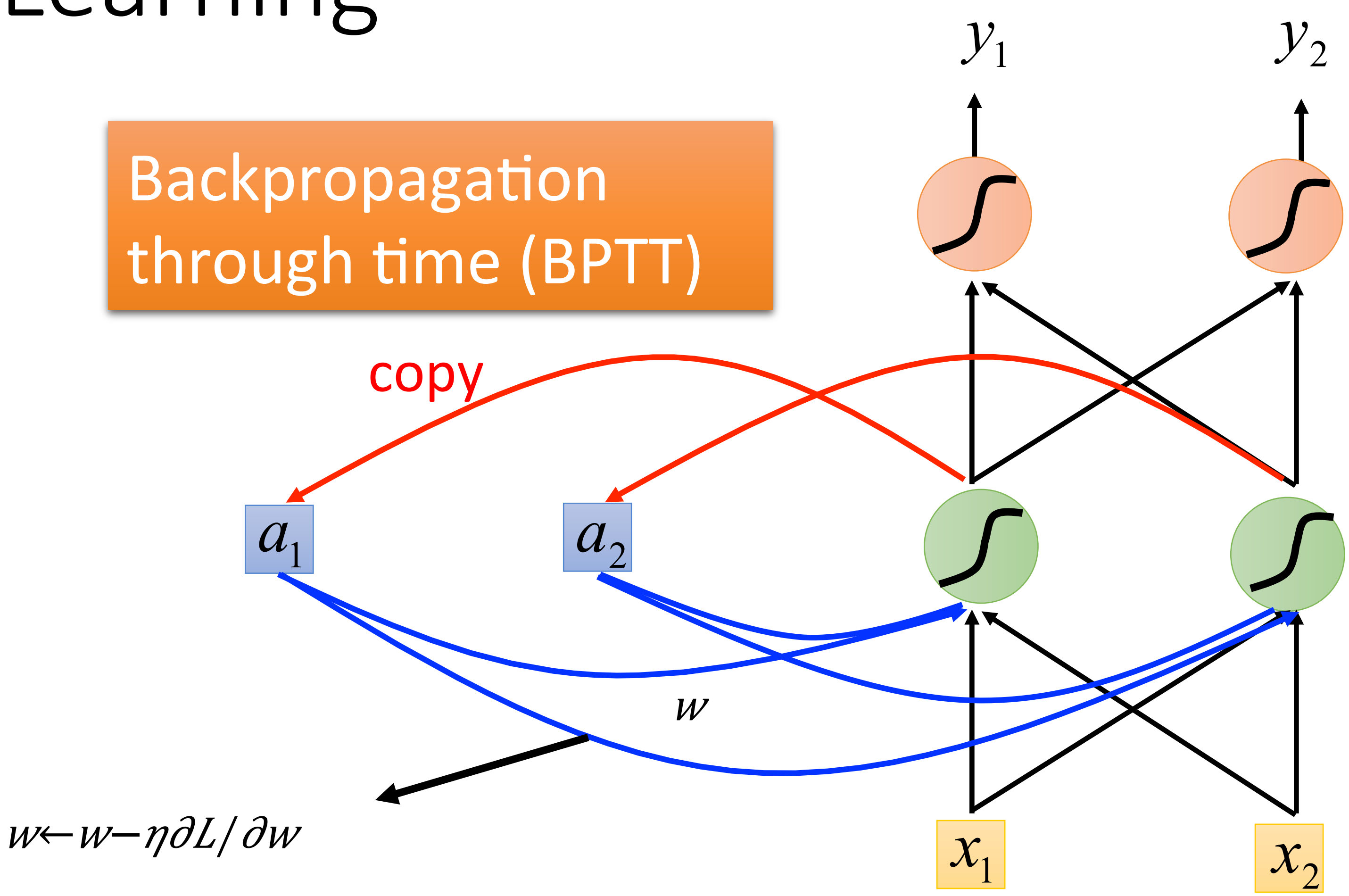
Step 2: Learning Target

Step 3: Learn!

天生的腦

Learning

Backpropagation through time (BPTT)



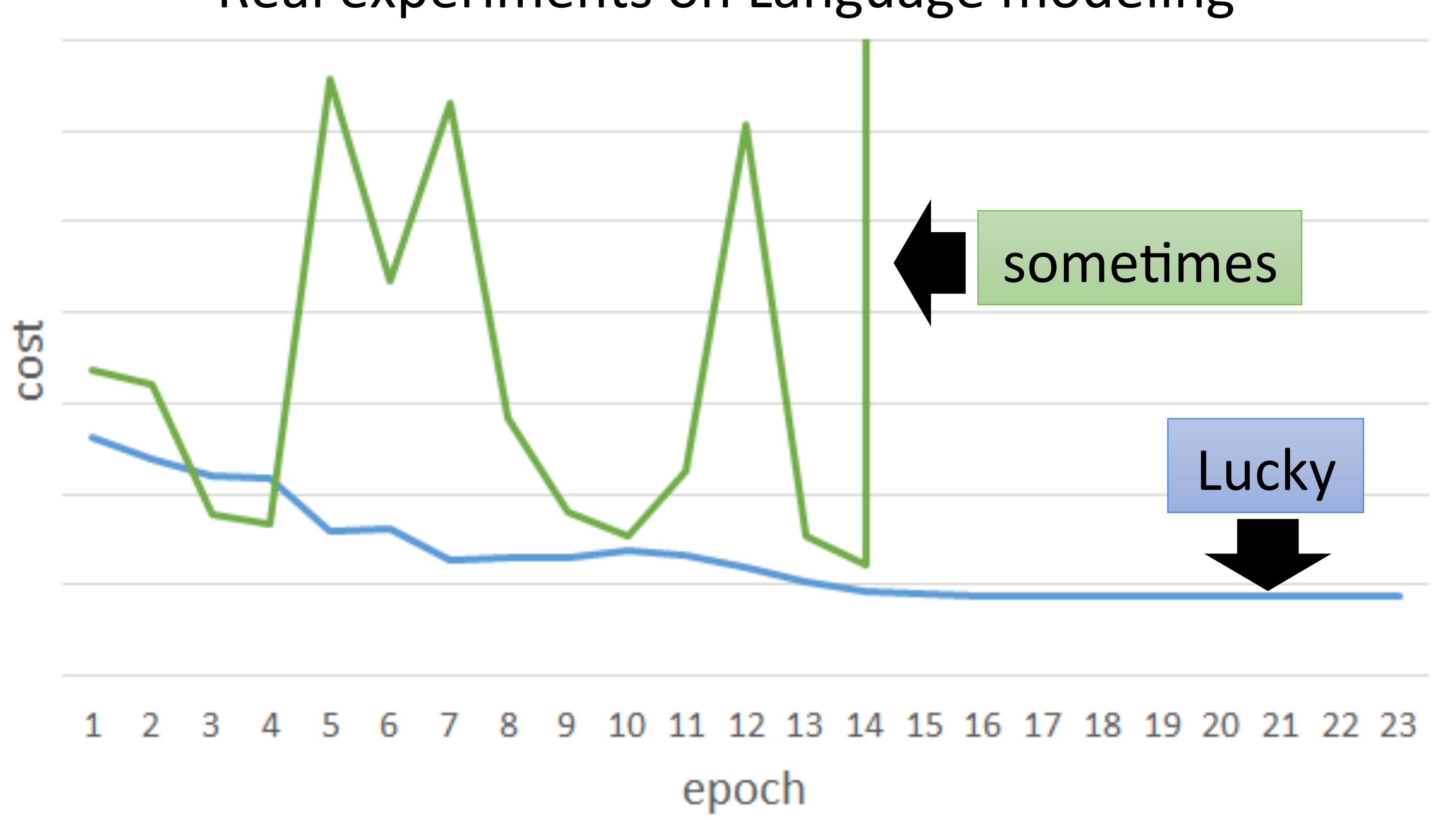
RNN Learning is very difficult in practice.

Unfortunately ……

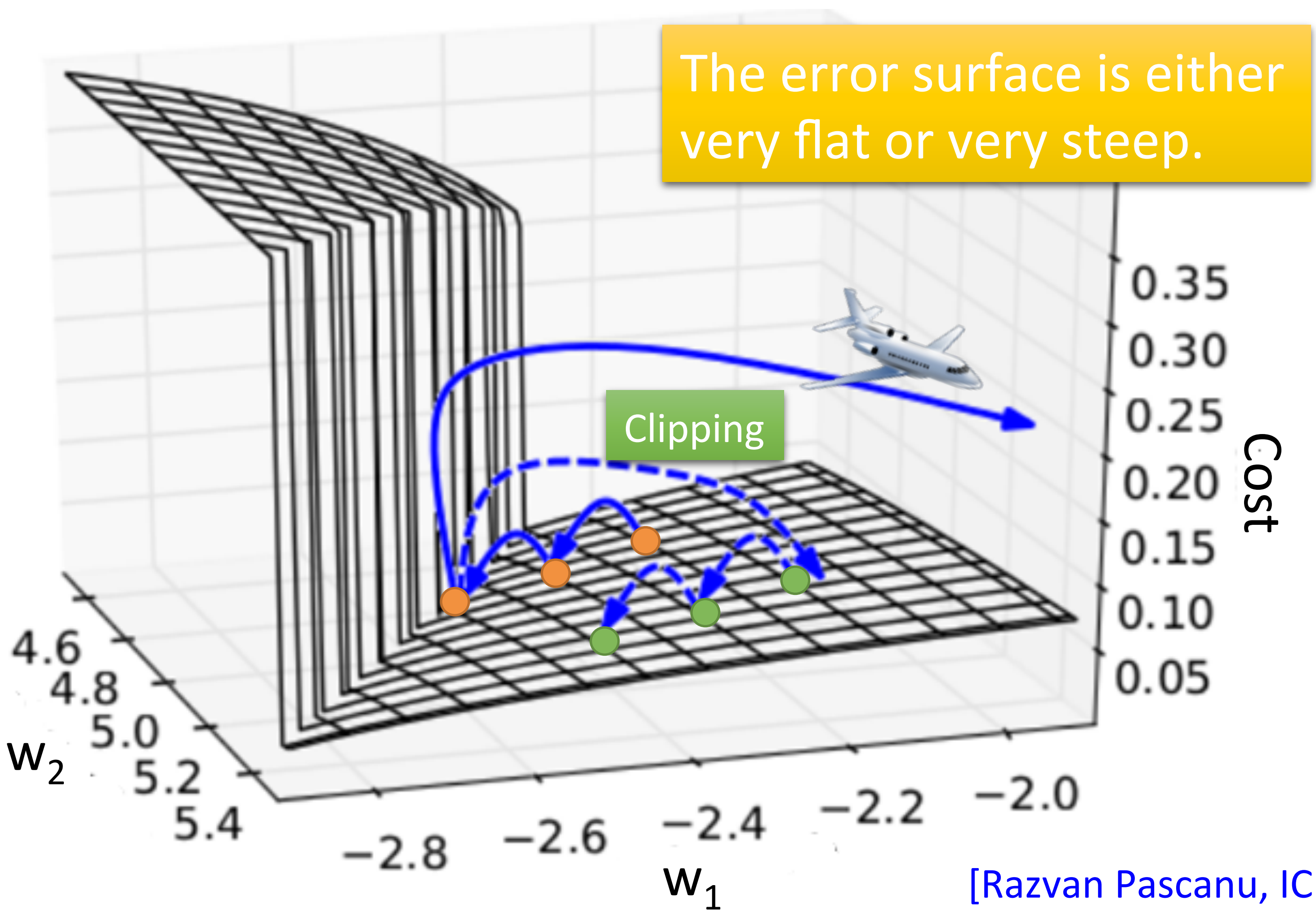
感謝 曾柏翔 同學
提供實驗結果

- RNN-based network is not always easy to learn

Real experiments on Language modeling



The error surface is rough.



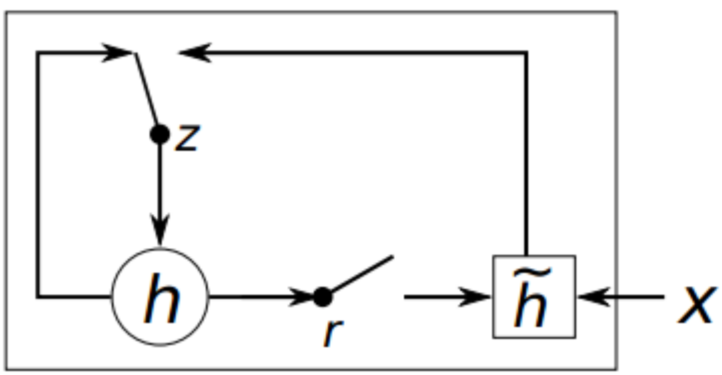
[Razvan Pascanu, ICML'13]

Helpful Techniques

- Nesterov's Accelerated Gradient (NAG):
  - Advance momentum method
- RMS Prop
  - Advanced approach to give each parameter different learning rates
  - Considering the change of Second derivatives
- Long Short-term Memory (LSTM)
  - Can deal with gradient vanishing (not gradient explode)

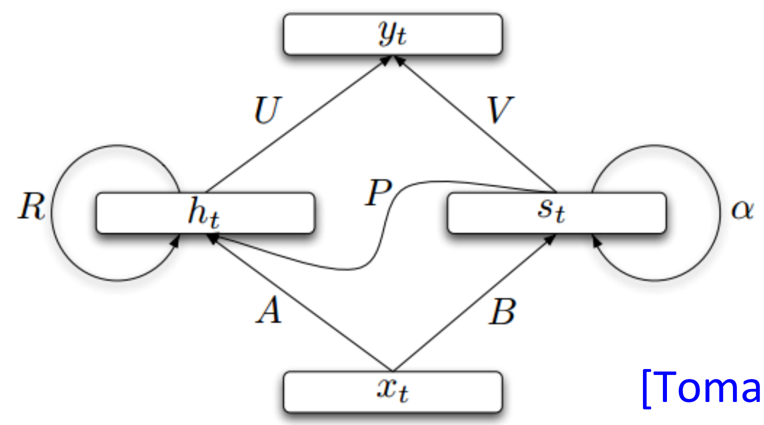
Helpful Techniques

Gated Recurrent Unit (GRU)



[Cho, EMNLP'14]

Structurally Constrained Recurrent Network (SCRN)



[Tomas Mikolov, ICLR'15]

Vanilla RNN Initialized with Identity matrix + ReLU activation function [Quoc V. Le, arXiv'15]

- Outperform or be comparable with LSTM in 4 different tasks

Outline of Lecture III

Recurrent Neural Network (RNN) & LSTM

More applications of RNN

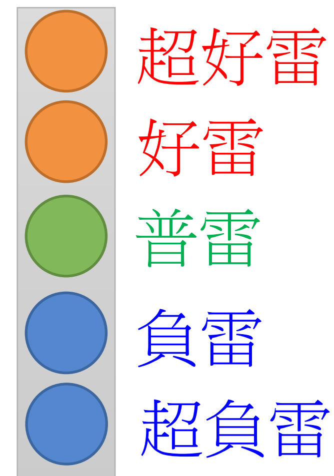
Next Wave: Attention–based Model

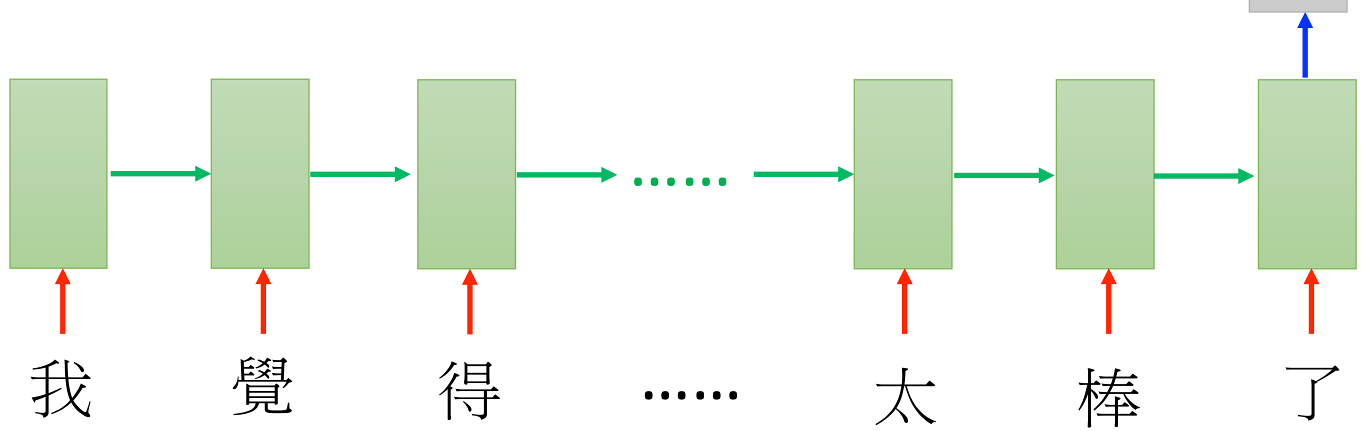
Many to one

- Input is a vector sequence, but output is only one vector

Sentiment Analysis

看了這部電影覺得很高興 …….	這部電影太糟了 …….	這部電影很棒 …….
Positive (正雷)	Negative (負雷)	Positive (正雷)



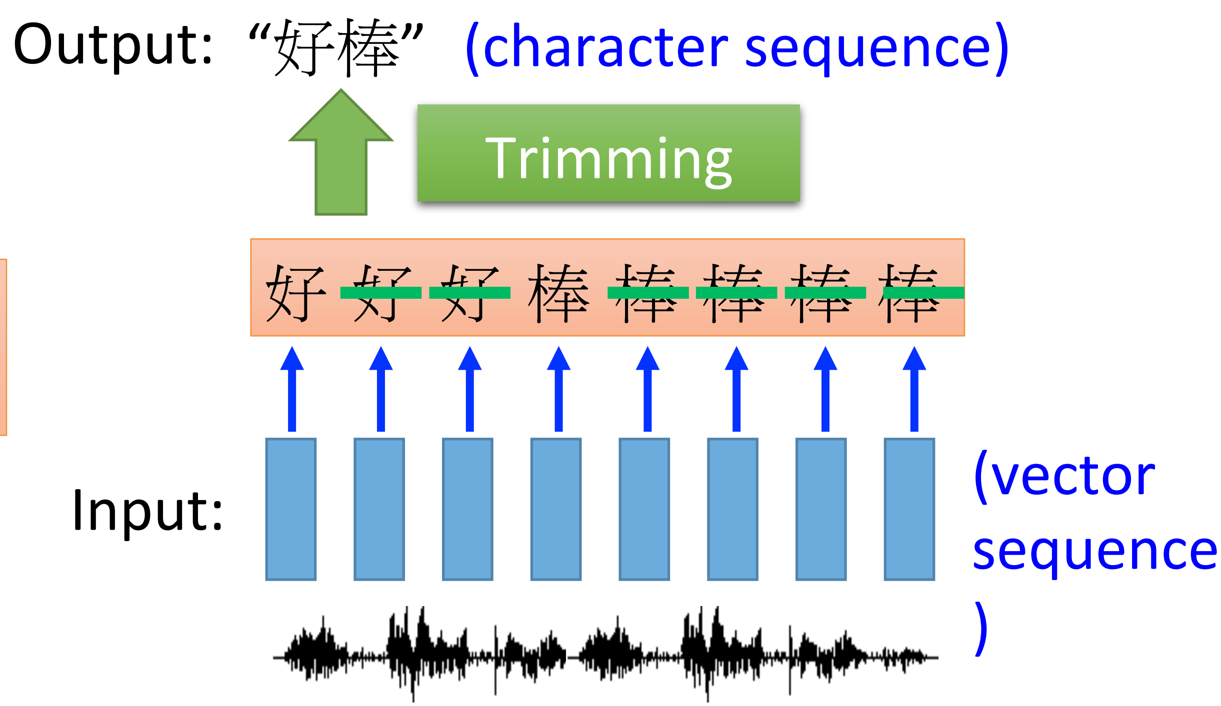


Many to Many (Output is shorter)

- Both input and output are both sequences, ***but the output is shorter.***
  - E.g. ***Speech Recognition***

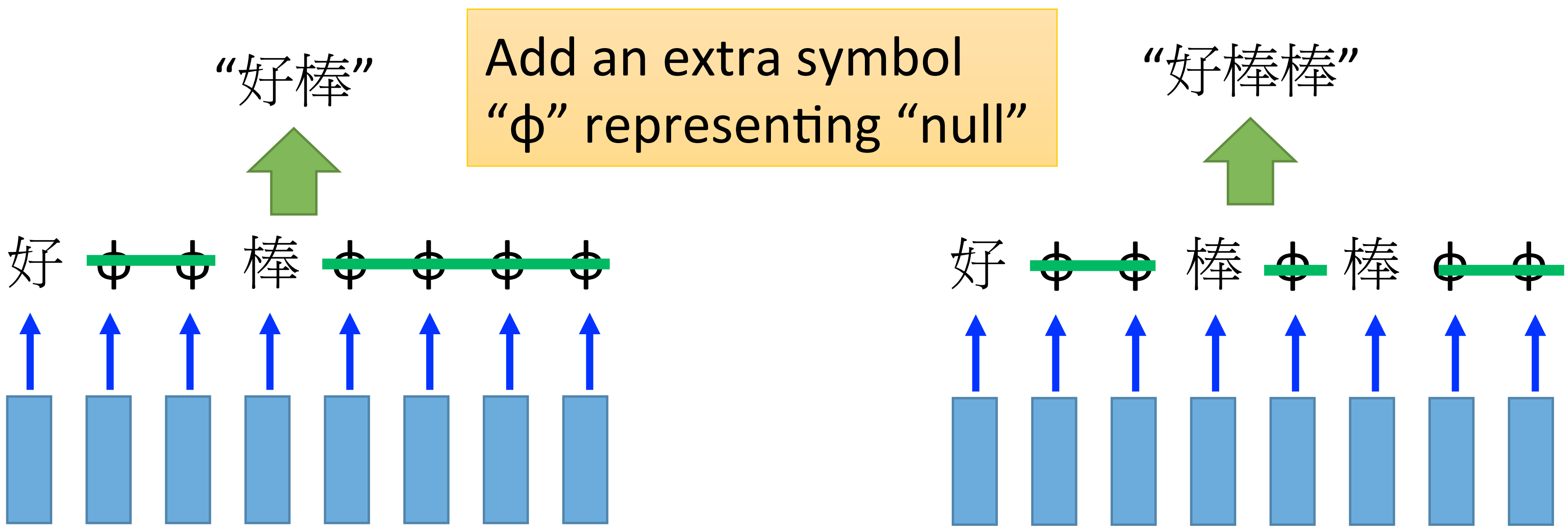
Problem?

Why can’t it be “好棒棒”



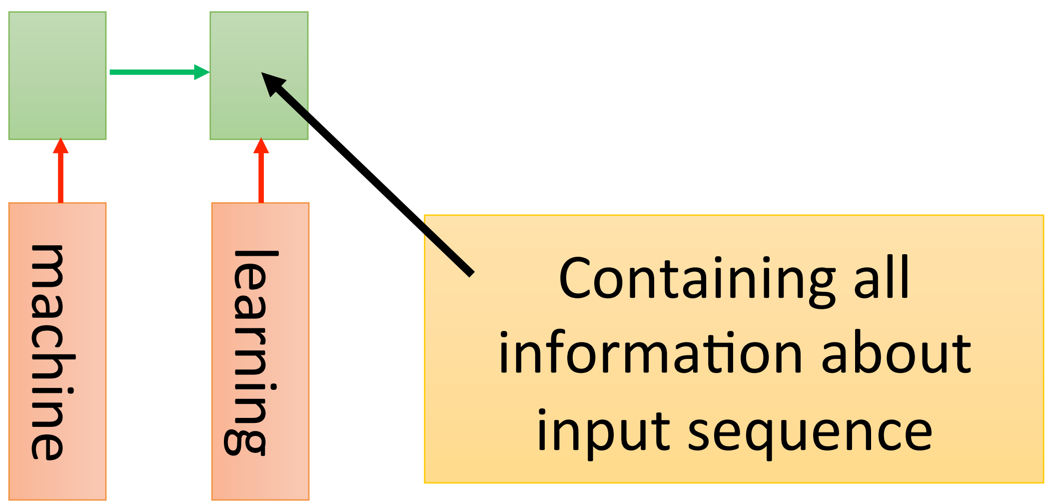
Many to Many (Output is shorter)

- Both input and output are both sequences, ***<u>but the output is shorter.</u>***
- Connectionist Temporal Classification (CTC) [Alex Graves, ICML'06][Alex Graves, ICML'14][Haşim Sak, Interspeech'15][Jie Li, Interspeech'15][Andrew Senior, ASRU'15]



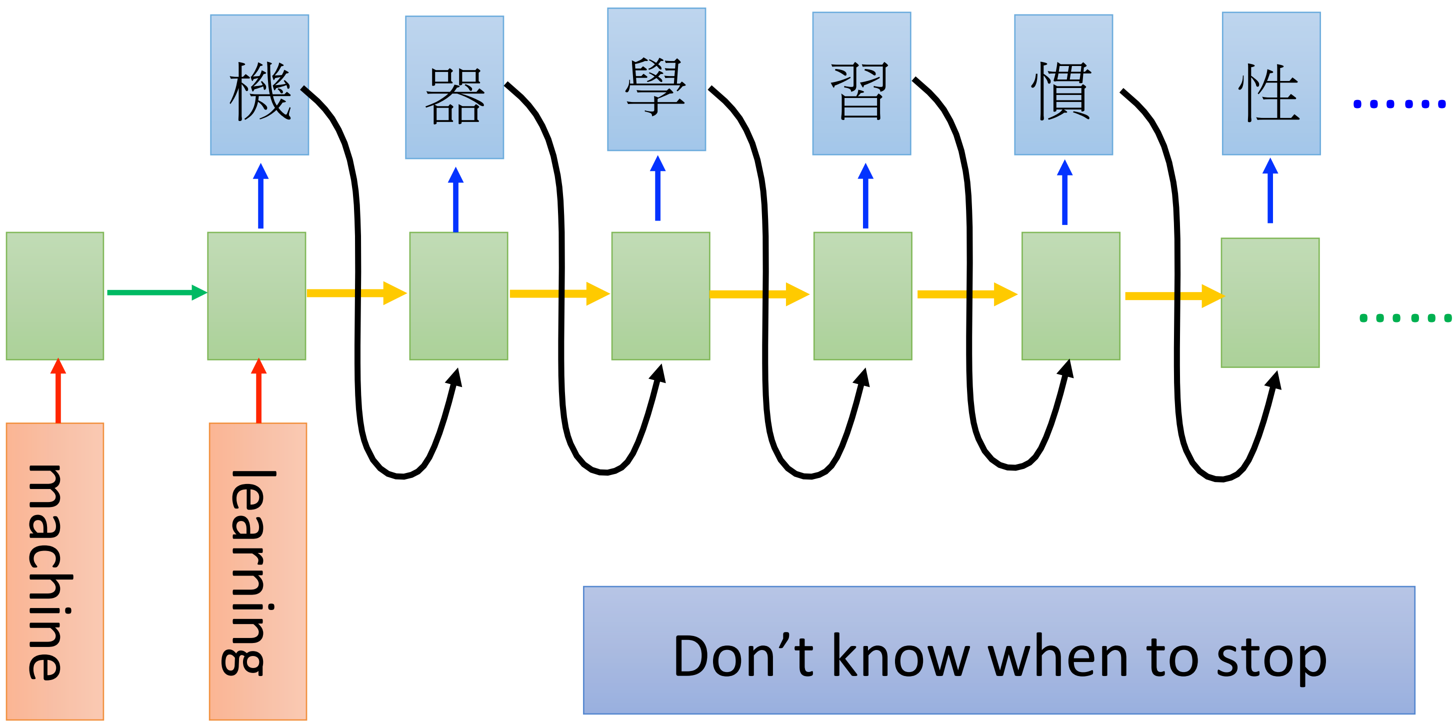
Many to Many (No Limitation)

- Both input and output are both sequences ***with different lengths***. → ***Sequence to sequence learning***
  - E.g. ***Machine Translation*** (machine learning→機器學習)

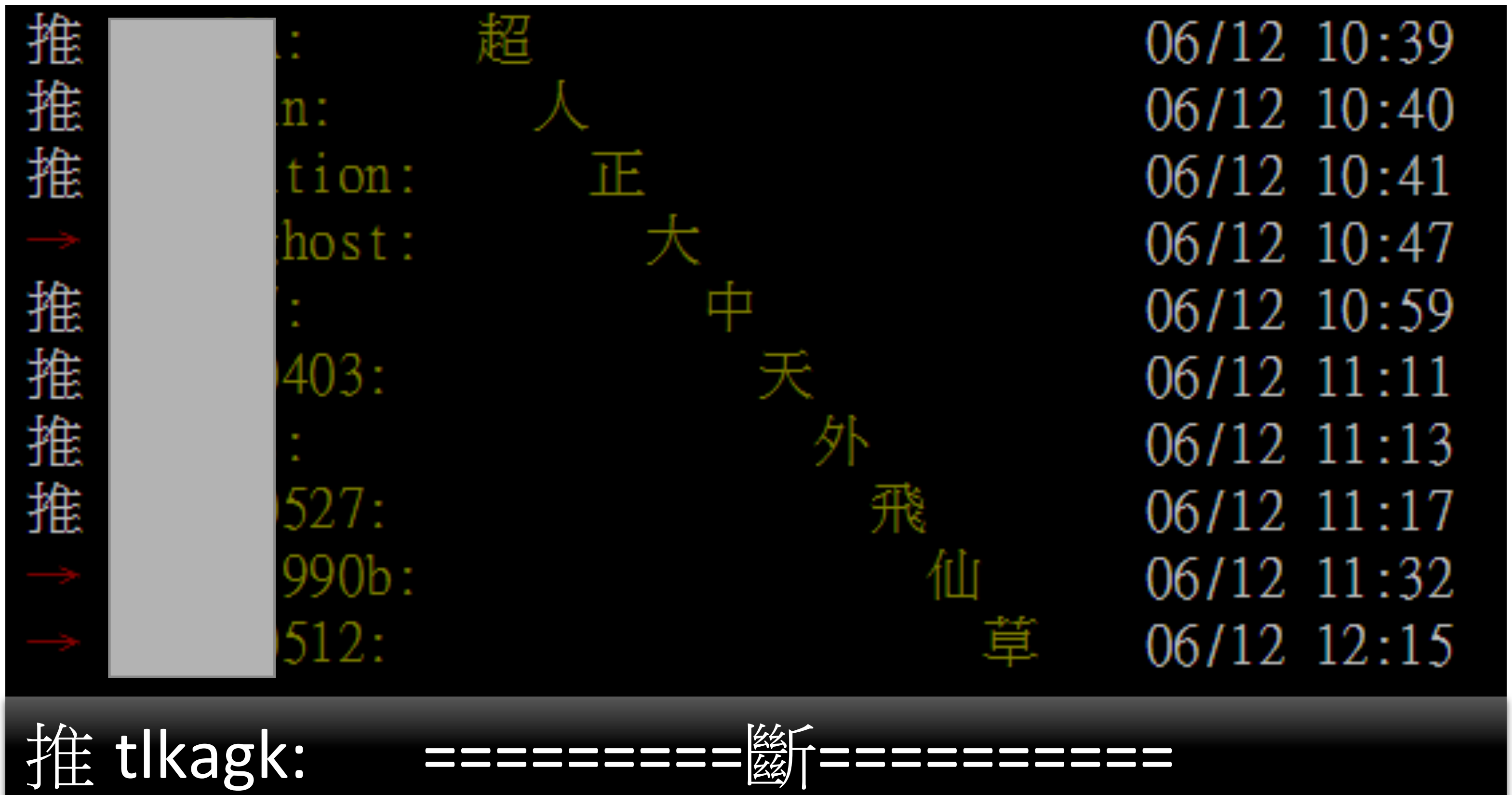


Many to Many (No Limitation)

- Both input and output are both sequences ***with different lengths***. → ***Sequence to sequence learning***
  - E.g. ***Machine Translation*** (machine learning→機器學習)



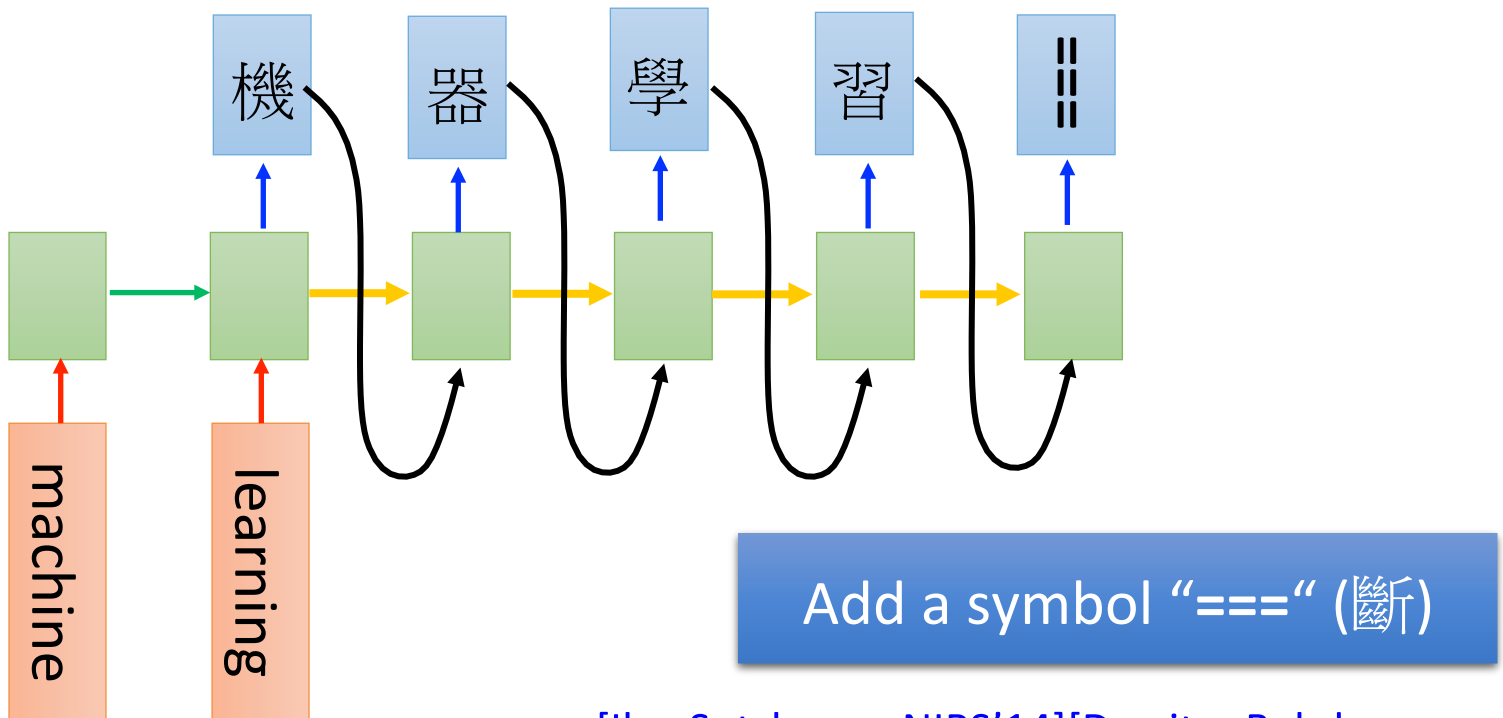
Many to Many (No Limitation)



Ref:http://zh.pttpedia.wikia.com/wiki/%E6%8E%A5%E9%BE%8D
%E6%8E%A8%E6%96%87 (鄉民百科)

Many to Many (No Limitation)

- Both input and output are both sequences ***with different lengths***. → ***Sequence to sequence learning***
  - E.g. ***Machine Translation*** (machine learning→機器學習)

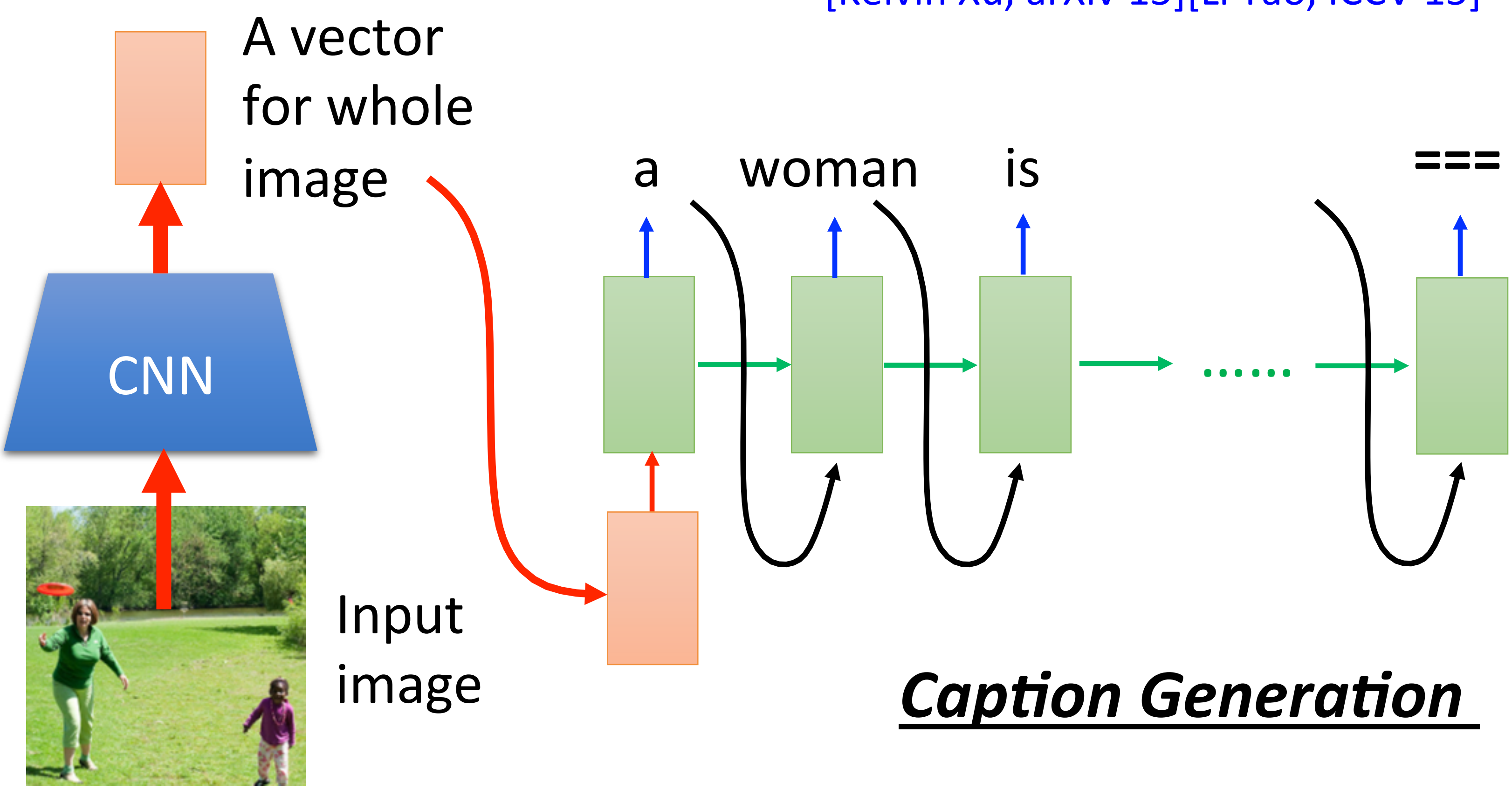


[Ilya Sutskever, NIPS'14][Dzmitry Bahdanau, arXiv'15]

One to Many

- Input an image, but output a sequence of words

[Kelvin Xu, arXiv'15][Li Yao, ICCV'15]



Caption Generation

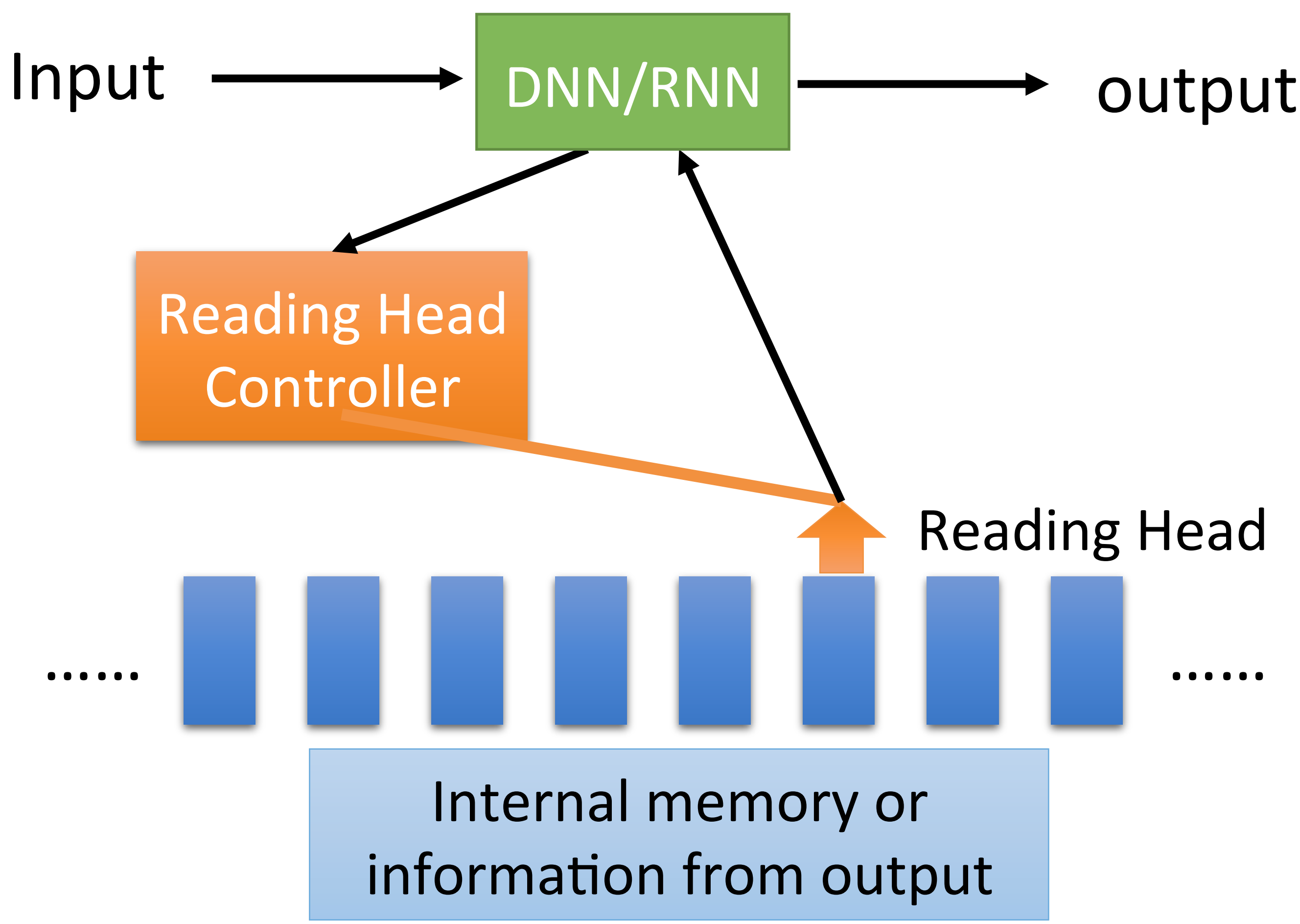
Outline of Lecture III

Recurrent Neural Network (RNN) & LSTM

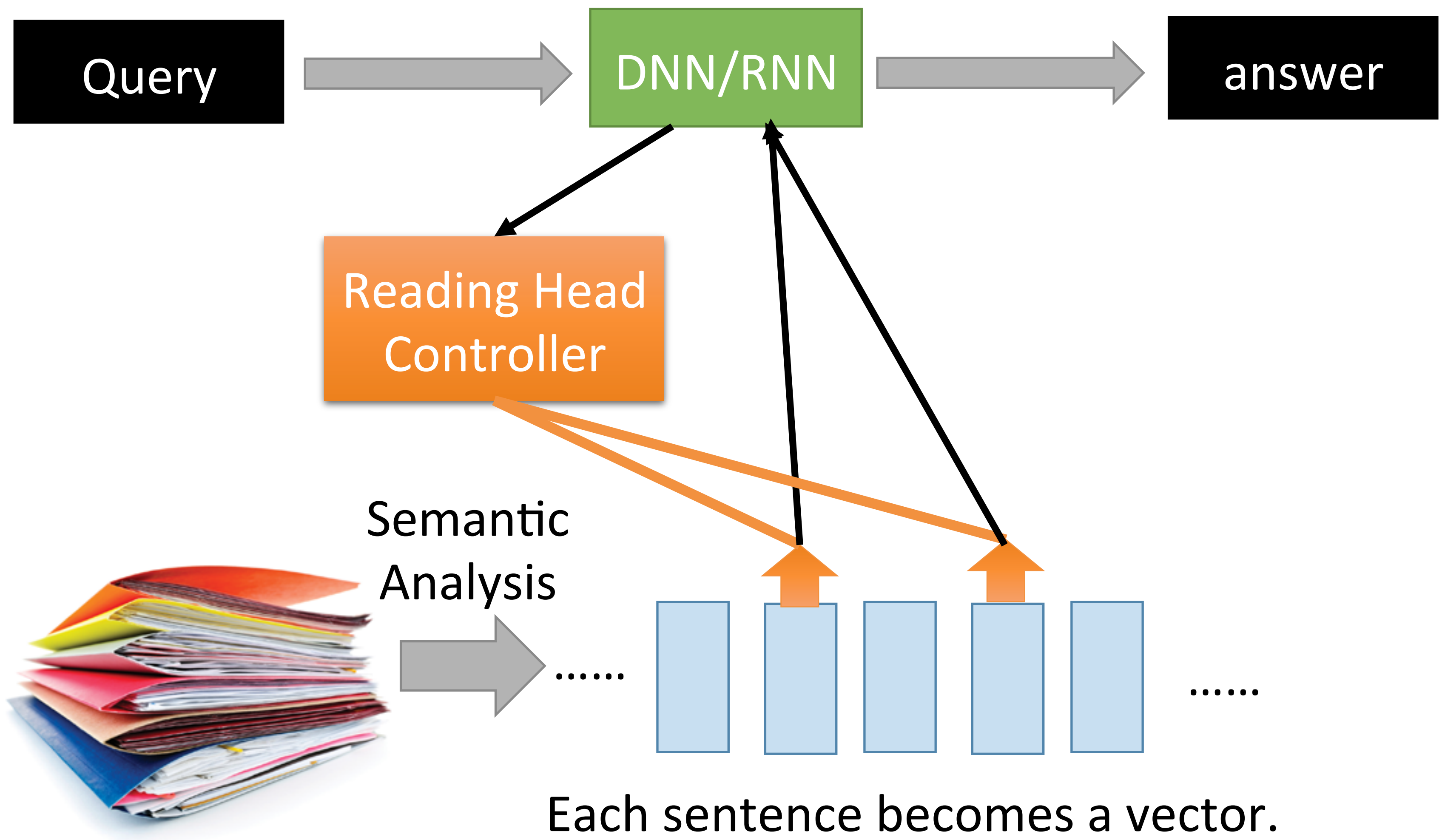
More applications of RNN

Next Wave: Attention–based Model

Attention-based Model



Reading Comprehension



Reading Comprehension

- End-To-End Memory Networks. S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus. NIPS, 2015.

The position of reading head:

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow Prediction: yellow				

Demo video: https://www.facebook.com/Engineering/videos/10153098860532200/

Visual Question Answering

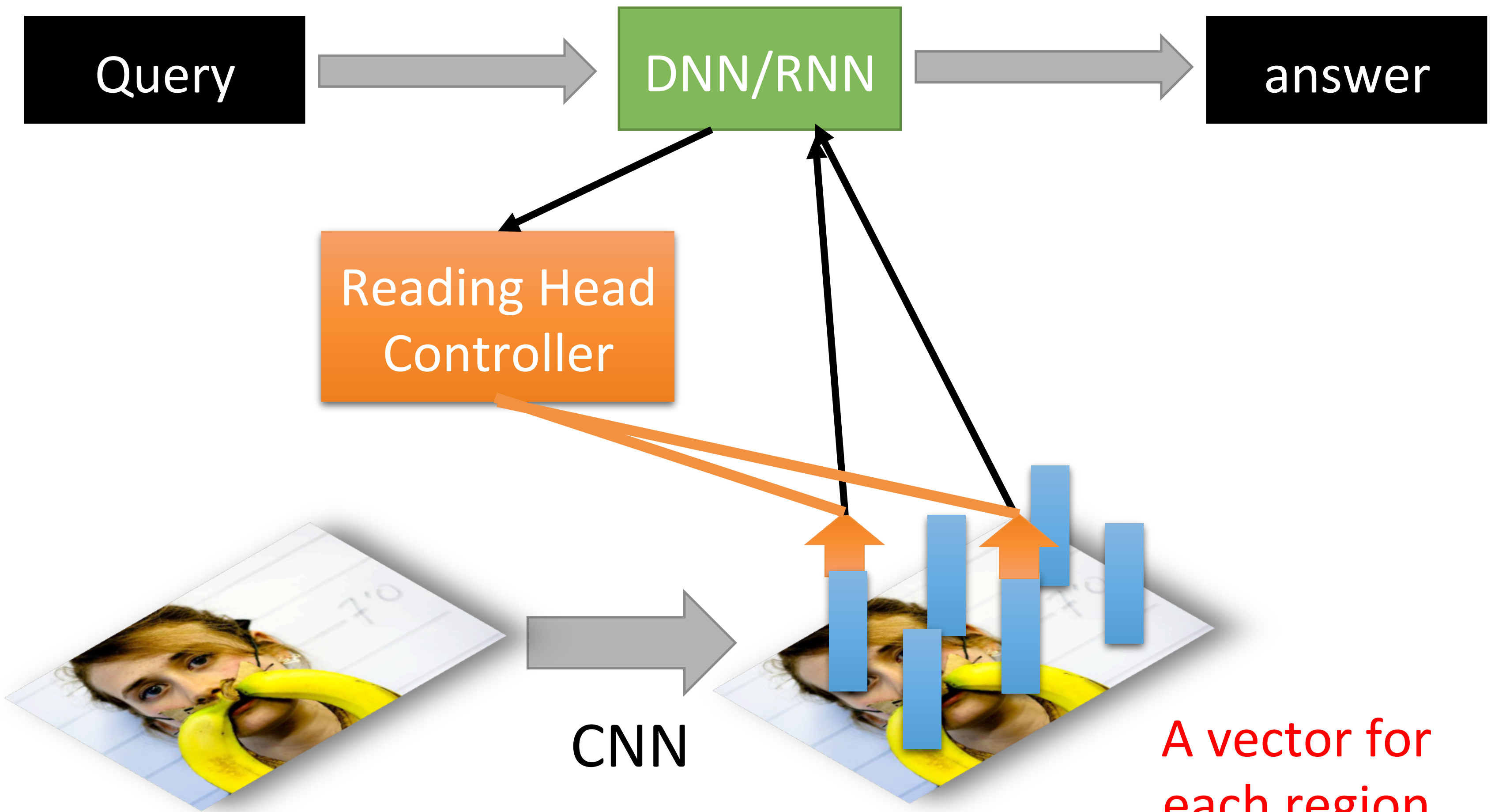
What is the mustache made of?

AI System

bananas

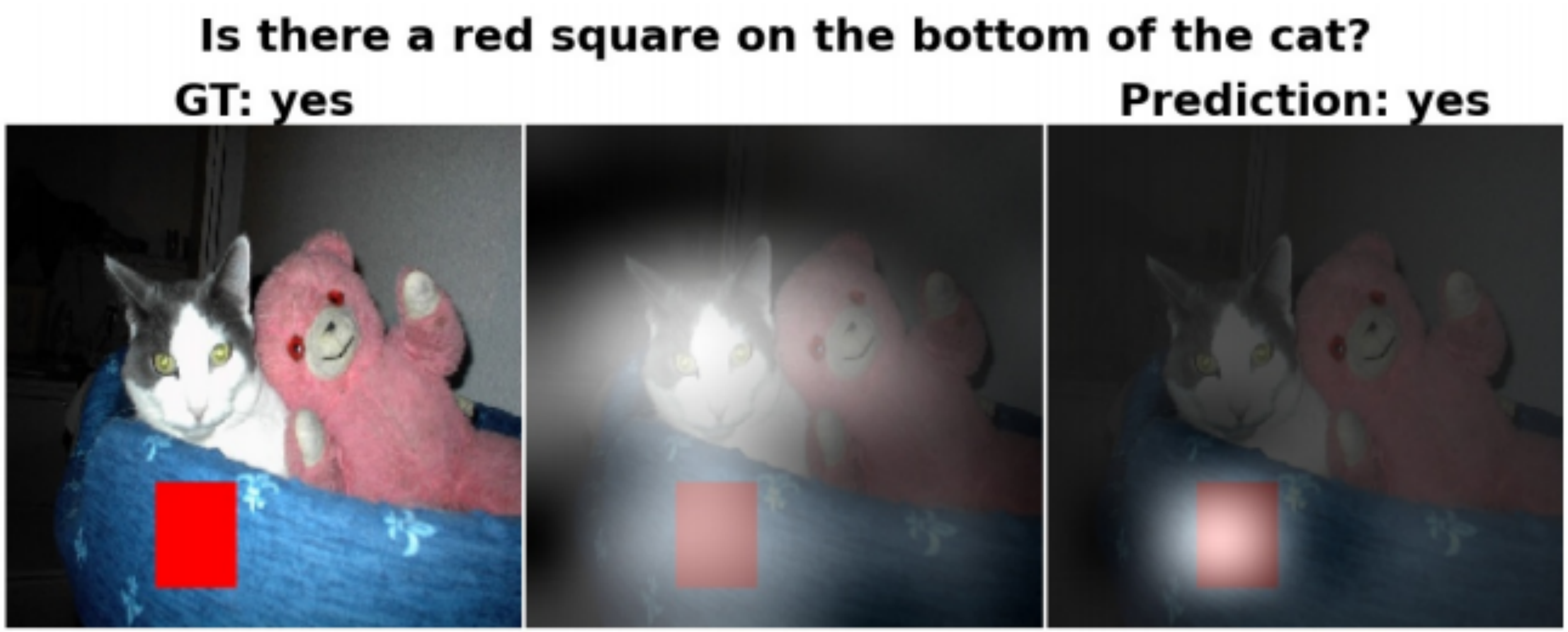
source: http://visualqa.org/

Visual Question Answering



Visual Question Answering

- Huijuan Xu, Kate Saenko. Ask, Attend and Answer: Exploring Question-Guided Spatial Attention for Visual Question Answering. arXiv Pre-Print, 2015



Homework 1

- Introduce a New NN with Memory

https://github.com/TheCEDL/homework1

Candidates

- Search RNN on Arxiv-sanity link
- Jianpeng Cheng et al. Long Short-Term Memory-Networks for Machine Reading. arXiv16'.
- Nal Kalchbrenner et al. Grid Long Short-Term Memory. arXiv16'. (From DeepMind, Alex)
- Kaisheng Yao et al. Depth-Gated LSTM. arXiv15'.
- Shuohang Wang et al. Learning Natural Language Inference with LSTM. arXiv15'.
- Junyoung Chung et al. Gated Feedback Recurrent Neural Networks. arXiv15'.