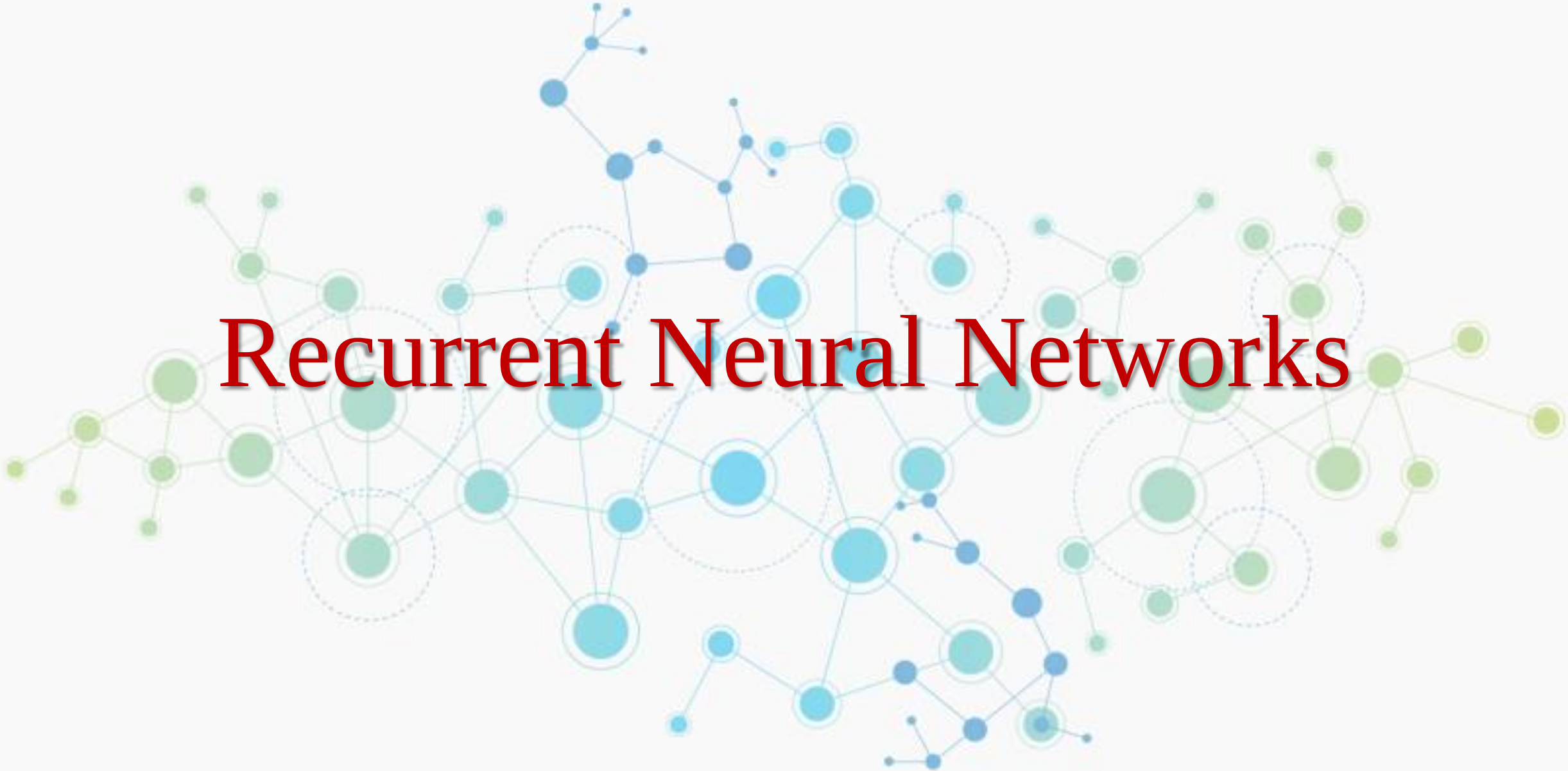
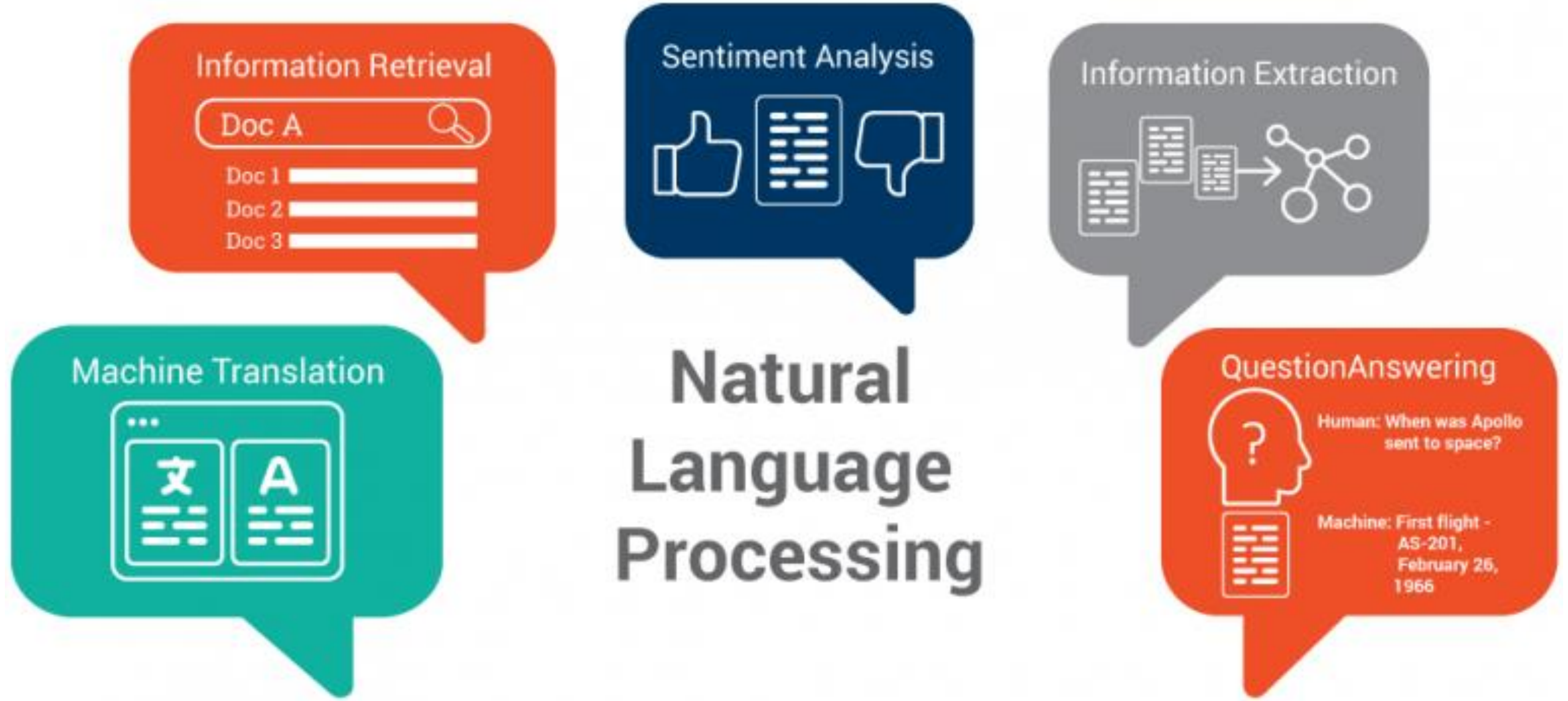
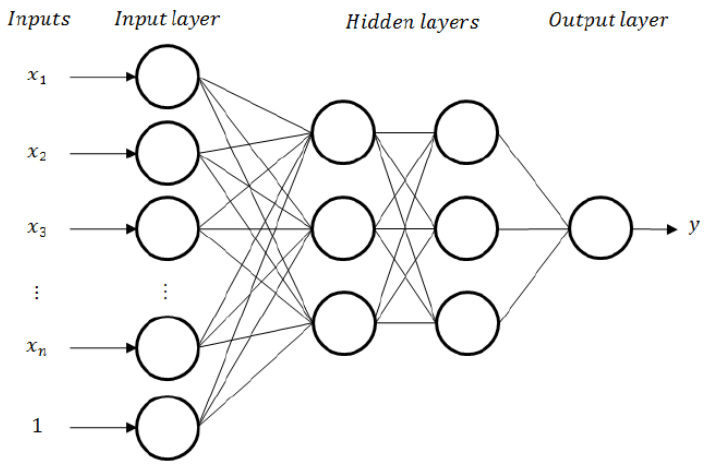


Recurrent Neural Networks


Information Retrieval
Doc A
Doc 1
Doc 2
Doc 3
Sentiment Analysis
Information Extraction
Machine Translation
文
A
Natural Language Processing
QuestionAnswering
Human: When was Apollo sent to space?
Machine: First flight - AS-201, February 26, 1966

Sentiment Analysis

- Let us try to classify following text as positive or negative
  - I like this phone – Positive
  - This phone is good – Positive
  - This phone is not okay – Negative
  - I do not like this phone because battery is not charging properly - Negative



Feed forward networks accept a fixed-sized vector as input !

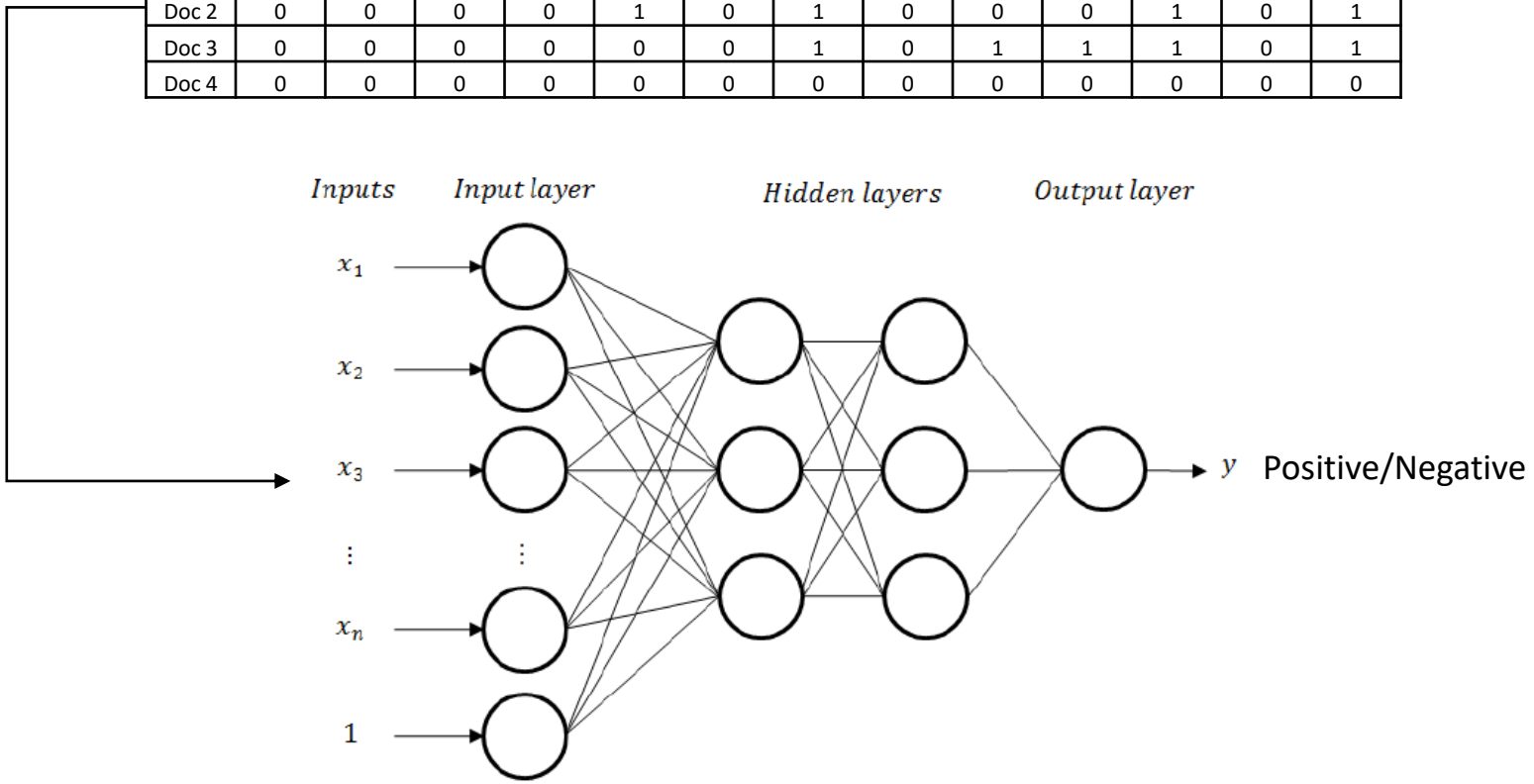
Solution 1: Using Bag-of-Words

- Represent the text using Bag-of-Words
  - I like this phone
  - This phone is good
  - This phone is not okay
  - I do not like this phone because battery is not good

	battery	because	charging	do	good	I	is	like	not	okay	phone	properly	this
Doc 1	0	0	0	0	0	1	0	1	0	1	0	0	1
Doc 2	0	0	0	0	1	0	1	0	0	0	1	0	1
Doc 3	0	0	0	0	0	0	1	0	1	1	1	0	1
Doc 4	0	0	0	0	0	0	0	0	0	0	0	0	0

Applying ANN

	battery	because	charging	do	good	I	is	like	not	okay	phone	properly	this
Doc 1	0	0	0	0	0	1	0	1	0	1	0	0	1
Doc 2	0	0	0	0	1	0	1	0	0	0	1	0	1
Doc 3	0	0	0	0	0	0	1	0	1	1	1	0	1
Doc 4	0	0	0	0	0	0	0	0	0	0	0	0	0



Drawback of BoW

- Lets try to represent the following text using Bag-of-Words
  - This phone is no good - Negative
  - No this phone is good - Positive

	good	is	no	phone	this
Doc 1	1	1	1	1	1
Doc 2	1	1	1	1	1

- Feed Forward Neural Networks with Bag-of-words (BoW) model does not consider position of words in input !

Word guessing game

- Guess the missing word in the follow sentence
  - I went to France last year, there the people speak the _____ language
- Lets find the missing word with the sentence (alphabetically sorted words)
  - Sorted Text: food I is it like Thai so
  - Original Text: I like Thai food it is so _____
  - Answer: I like Thai food it is so _____
- Feed Forward Neural Network will fail for guessing missed words
- Conclusion: Sequence matters !

Other situations where sequence matters

- Stock price today will be more or less similar to yesterday's price
- Tomorrow's temperature will be close to todays' temperature

Sequence Application Variation

- Audio Signal to Sequence - Speech Recognition
- Nothing to Sequence or Single Parameter to Sequence - Music Generation
- Sequence to Single Output - Sentiment Classification
- Sequence to Sequence - Machine Translation
- Video Frame Sequence to Output - Activity Recognition
- Sub-Sequence from a Sequence - Finding Specific Protein from a DNA Sequence
- Outlining Specific parts of a sequence - Name Entity Recognition

Notation Understanding

X: Rama Conquered Ravana to install the virtue of dharma

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

$T_x = 9$ (Length of training sequence: 9)

$x^{i<t>}$: t^{th} word of i^{th} training sequence

Y: 1 0 1 0 …… 0 0 0 0

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

$T_y = 9$ (Length of output sequence: 9)

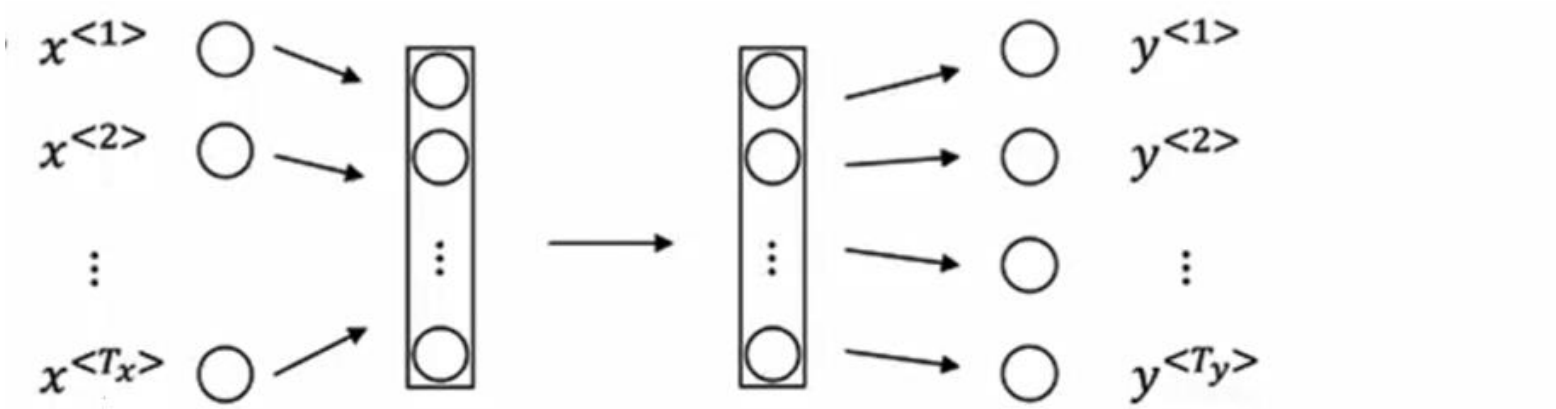
$y^{i<t>}$: t^{th} word of i^{th} output sequence

Representing words and one-hot encoding

X: Rama Conquered Ravana to install the virtue of dharma

A	1	Rama	Ravana
:		0	0
:		0	0
Conquered	329	0	0
:		0	0
:		0	0
Install	4521	:	:
:		:	:
:		:	:
Rama	7689	1 -7689	:
:		:	1-7900
Ravana	7900	:	:
:		0	0
ZZZ	10000	0	0

Standard Neural Network Does not works out to give a good application for sequence models

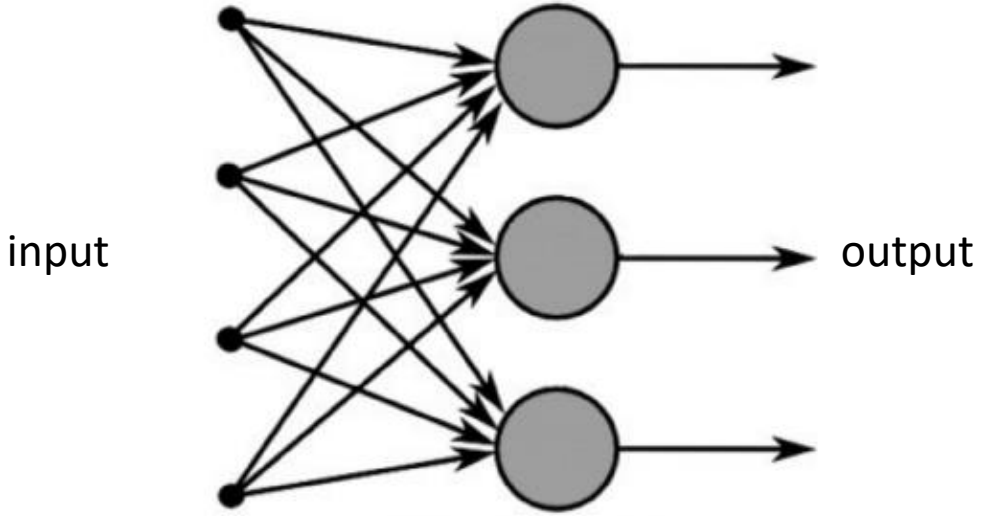


Inputs, outputs can be different lengths in different examples.

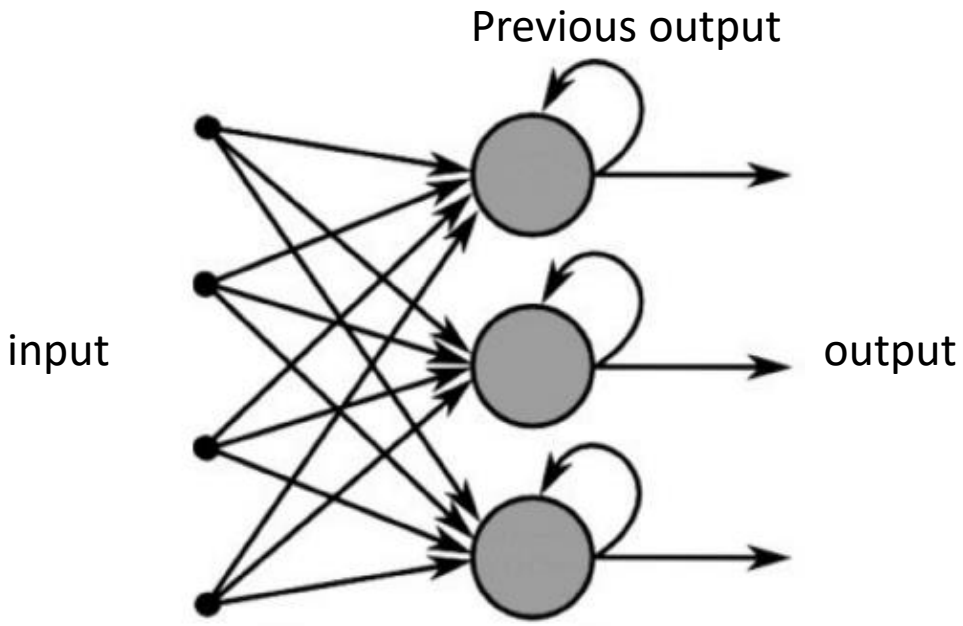
Doesn't share features learned across different positions of text.

- Feed forward networks accept a fixed-sized vector as input and produce a fixed-sized vector as output
- So, Feed forward networks cannot process sequential data containing variable length of data
- Feed forward networks does not consider sequence in the data

NN vs RNN

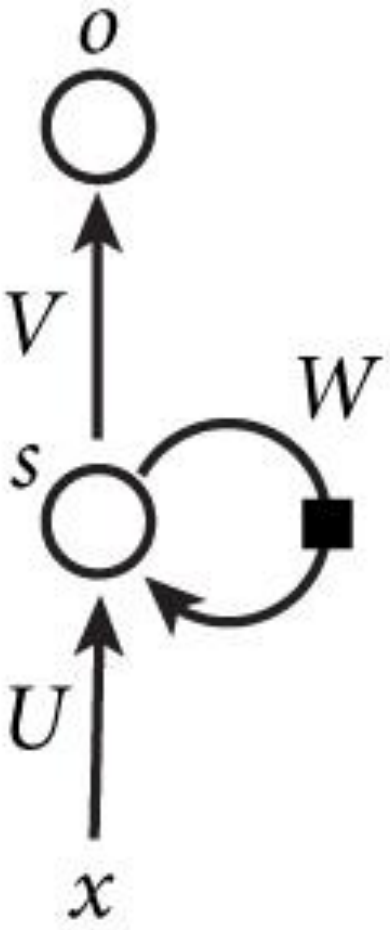


Feed-Forward Neural Network



Recurrent Neural Network

Notations



x : Input

o : Output

s : state of the hidden unit

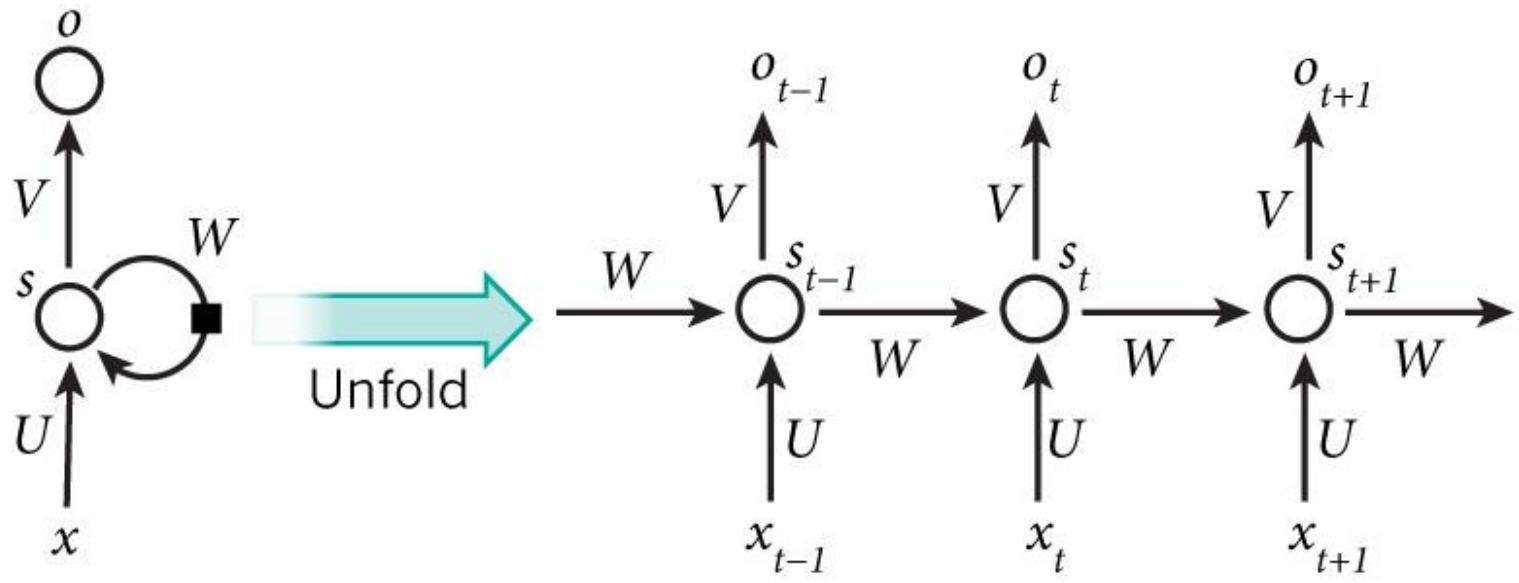
U, V and W : Weights to be learned

U : weights used for hidden state computation (from input)

V : weights used for output computation

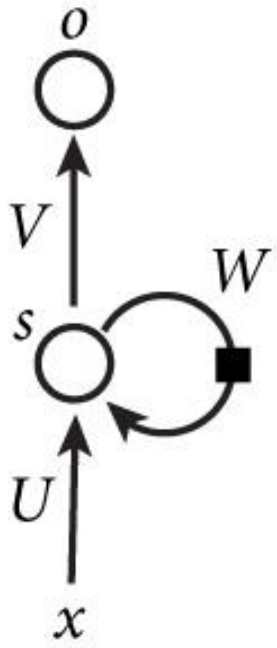
W : weights used for hidden state computation (from previous hidden state)

Unrolled RNN with parameters



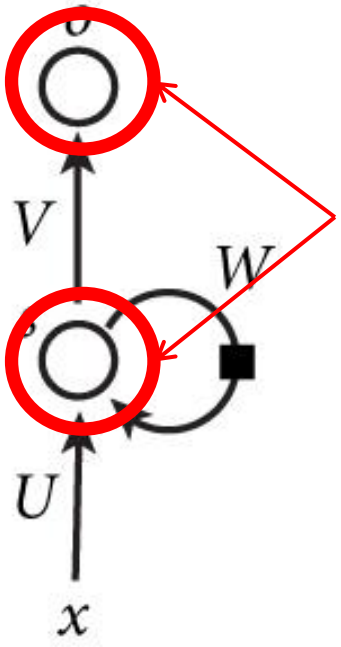
The recurrent network can be converted into a feed forward network by **unfolding over time**

RNN Forward Pass

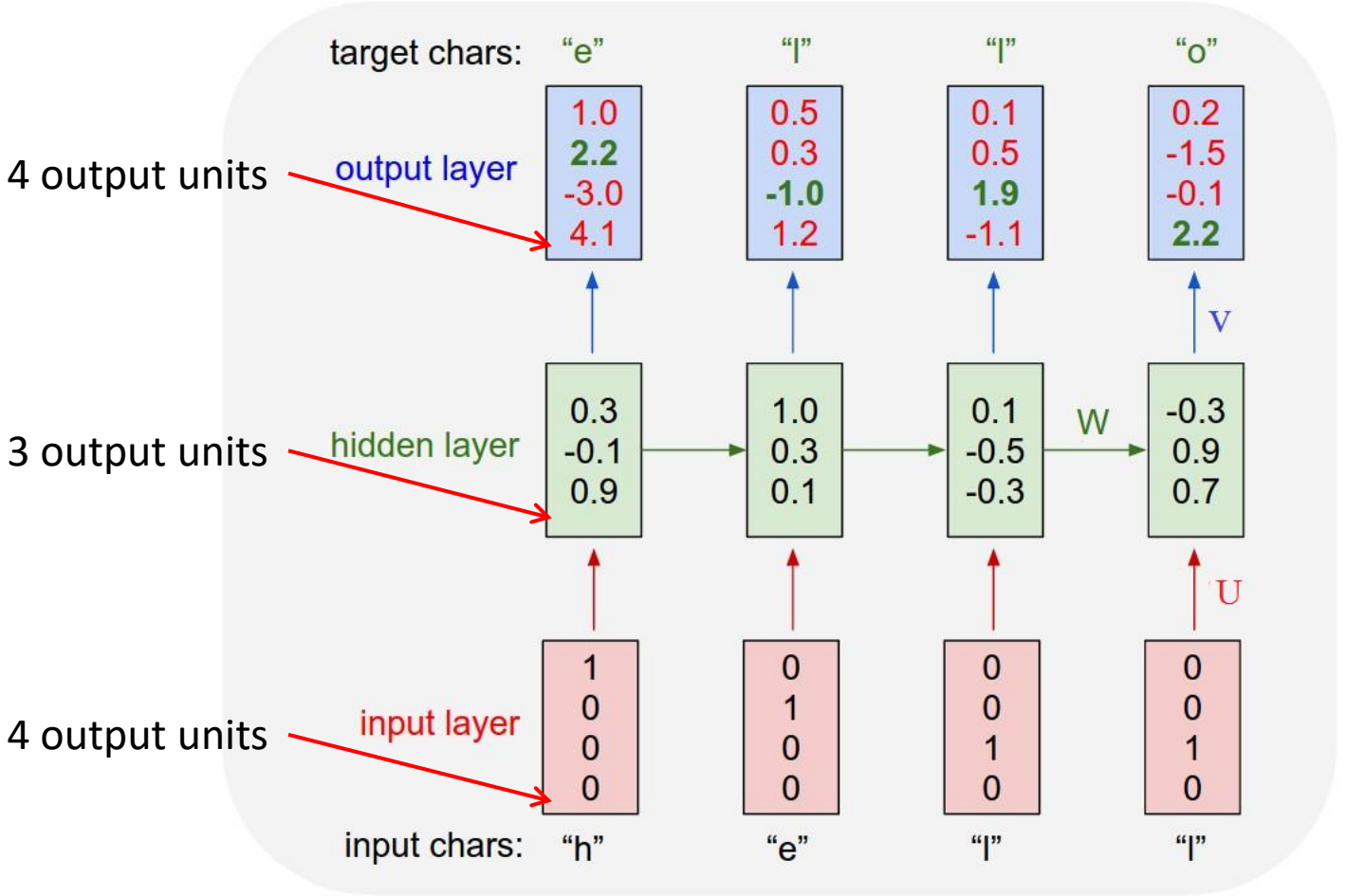


- Step 2: Current hidden state *s* at time *t* will be computed using
$$s(t) = f_h(Ux(t) + Ws(t\ -1)\ )$$
- Step 1: input *x* will be given at time *t*
- Step 3: Current output *o* at time *t* will be computed using
$$o(t) = f_o\big(Vs(t)\big)$$
- Note: Output will not necessarily be generated for every *t*. i.e. it depends upon application. Speech recognition RNN will output words instantly at every iteration. Opinion classification RNN will output label only at the end of sentence.

Important Notes



These are not single units
There are single layers

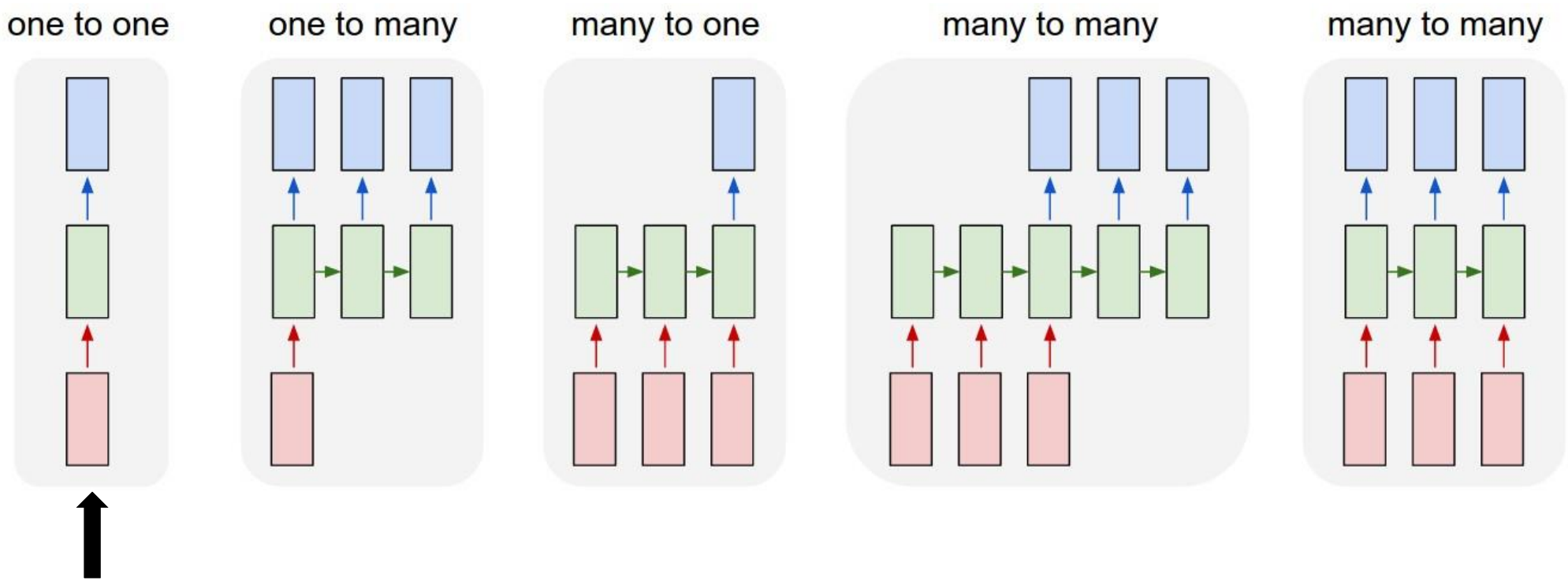


RNN using Keras

- **</>**

- keras.layers.RNN(cell, return_sequences=False, return_state=False, go_backwards=False, stateful=False, unroll=False)

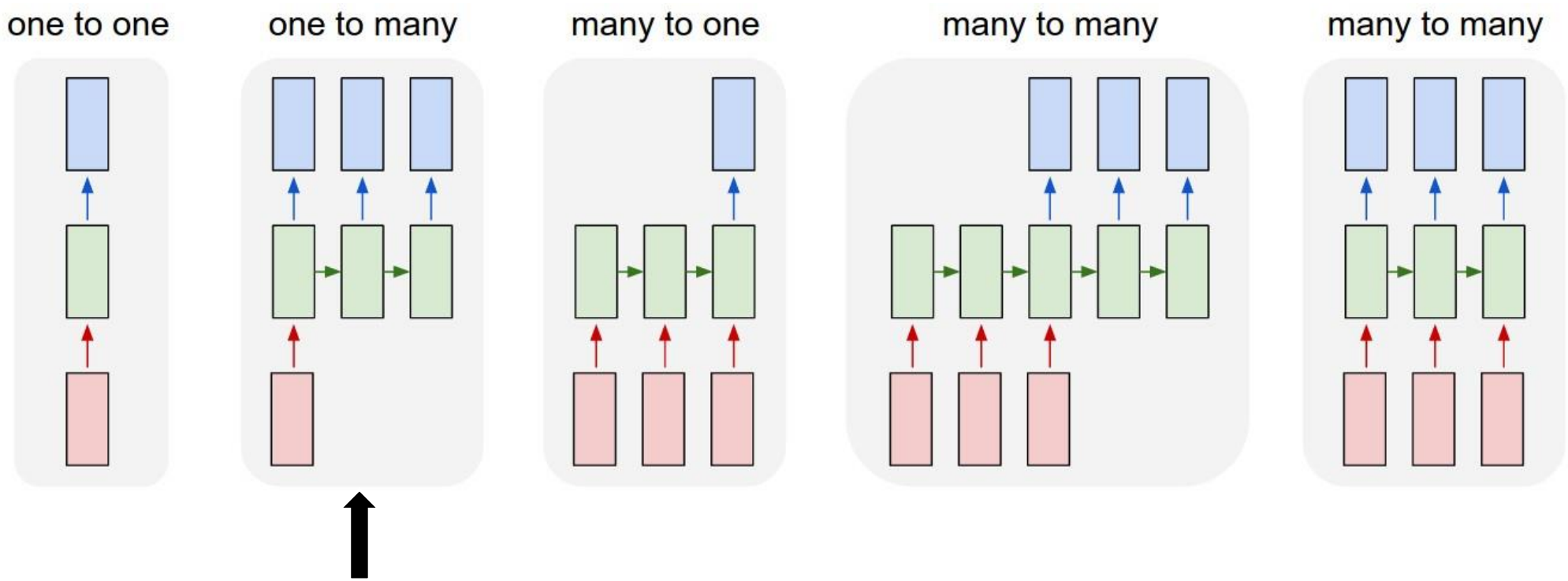
RNN Variants



Vanilla mode of processing without RNN, from fixed-sized input to fixed-sized output (e.g. image classification)

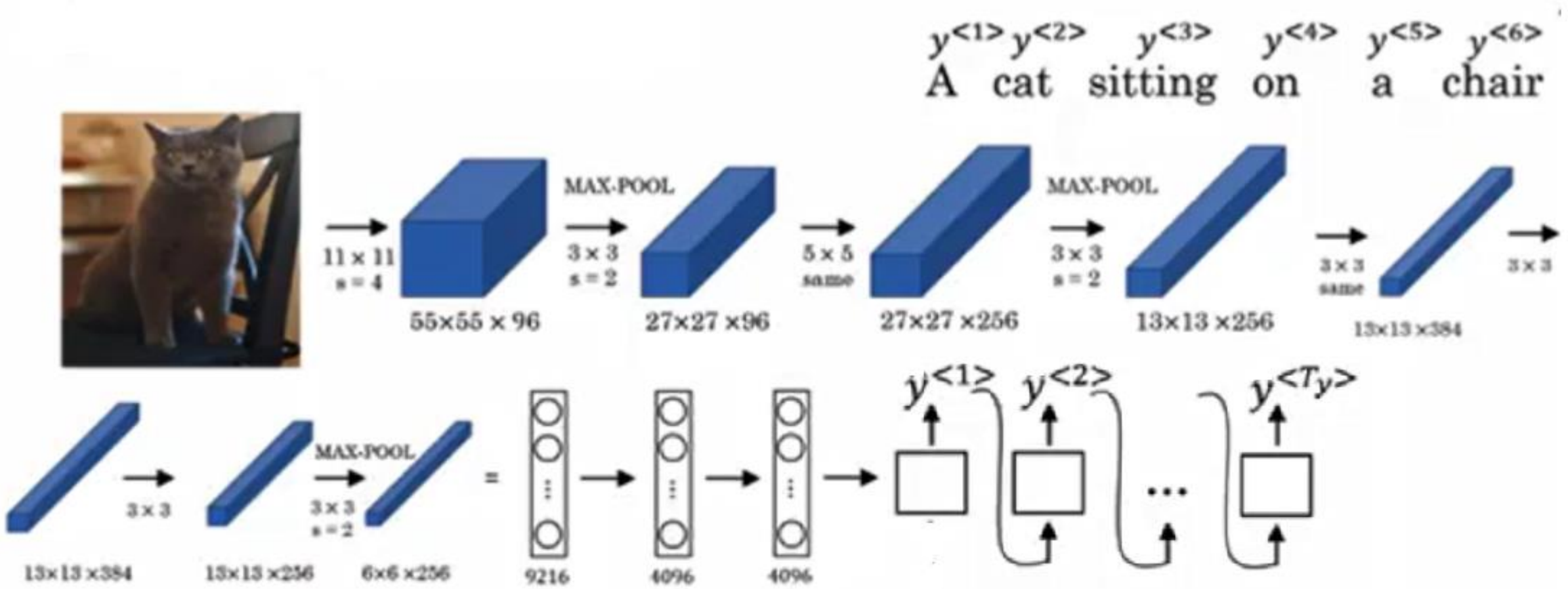
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

RNN Variants

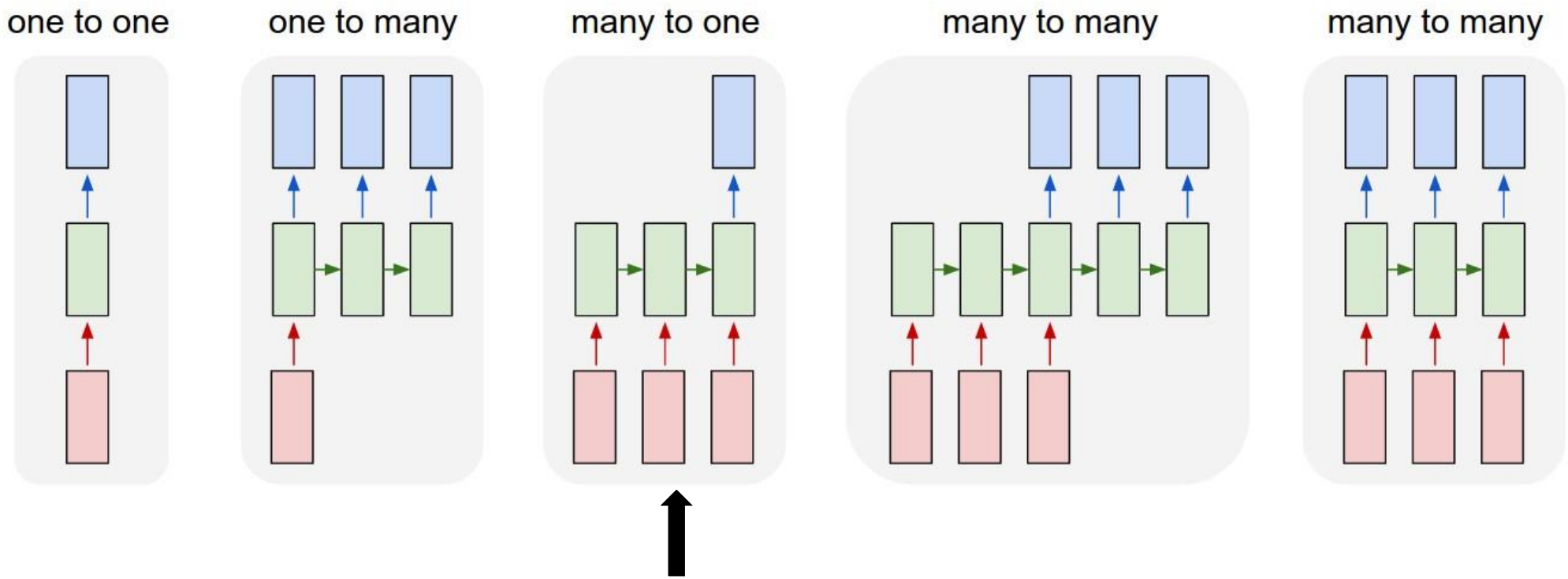


Sequence output (e.g. image captioning takes an image and outputs a sentence of words)

Image captioning model

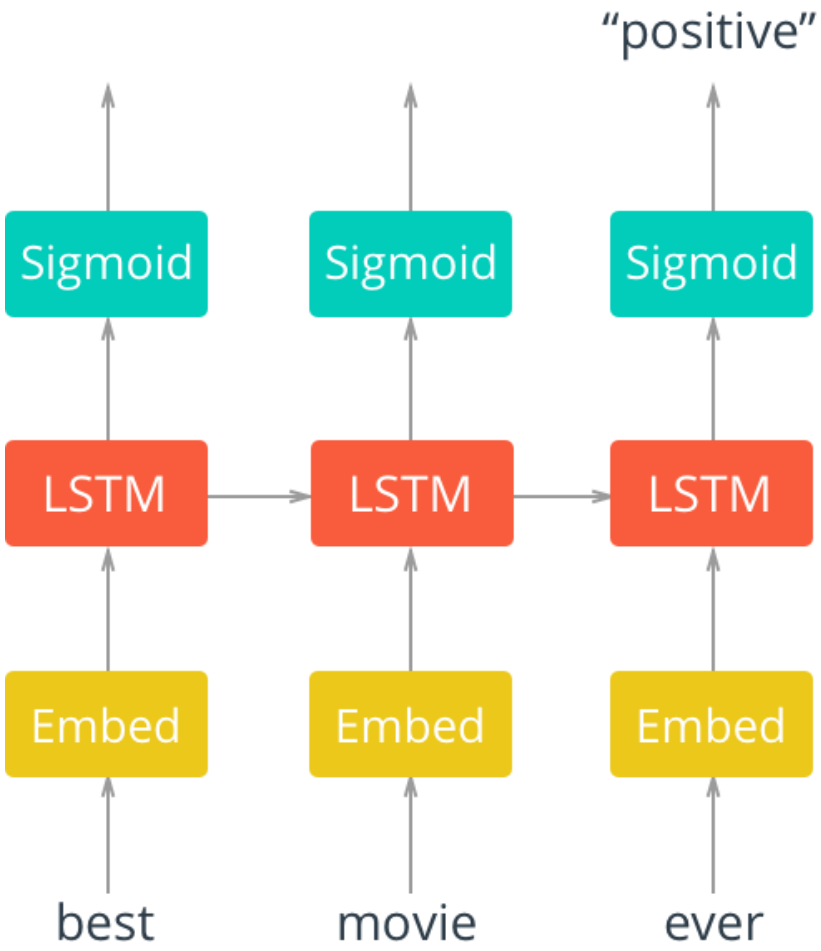


RNN Variants

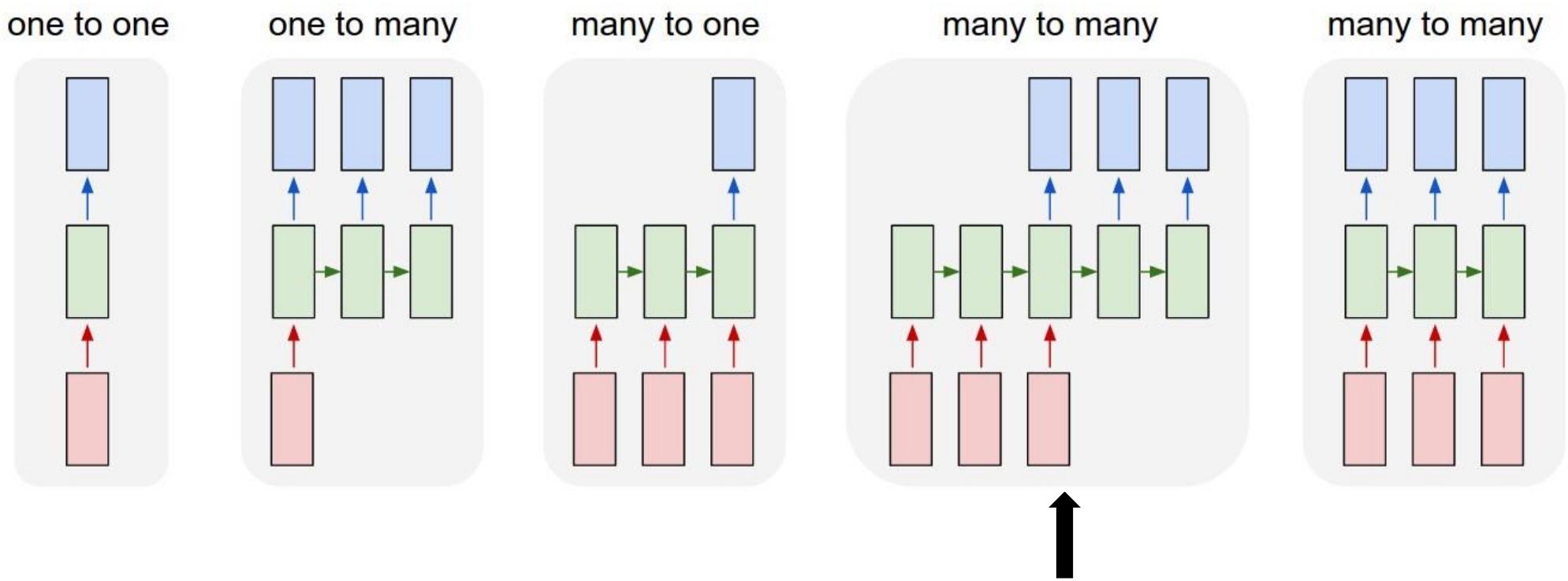


Sequence input (e.g. sentiment analysis where a given sentence is classified as expressing positive or negative sentiment)

Sentiment Classification

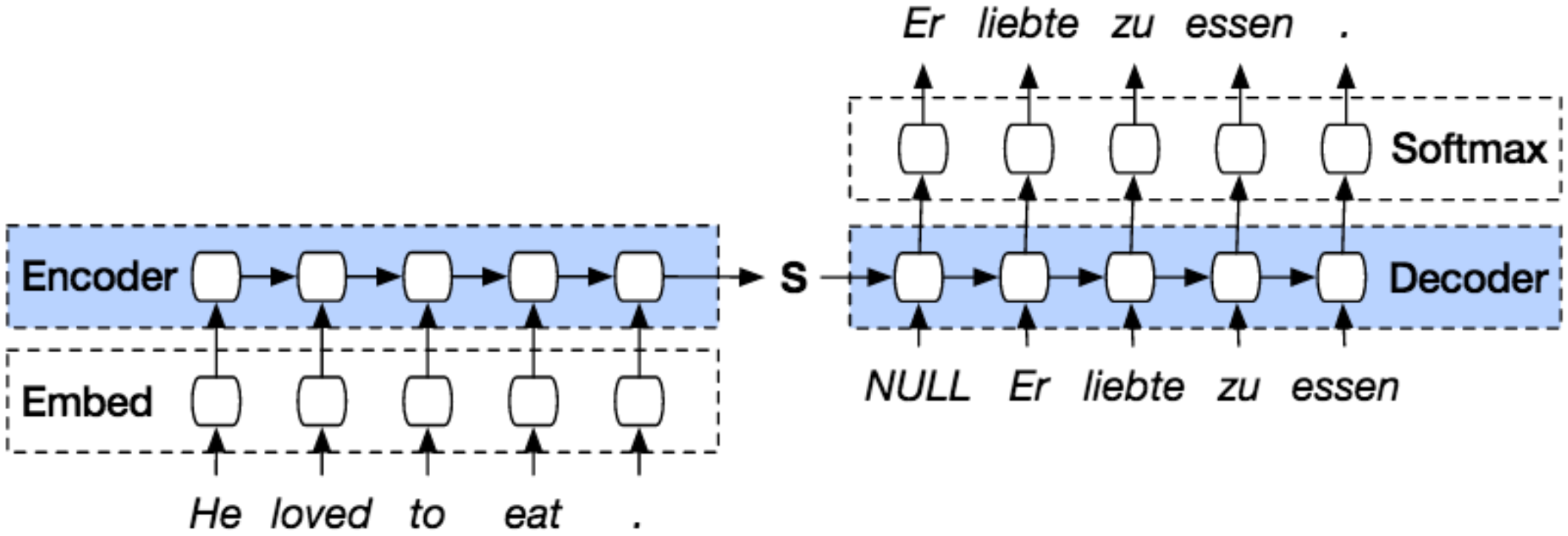


RNN Variants

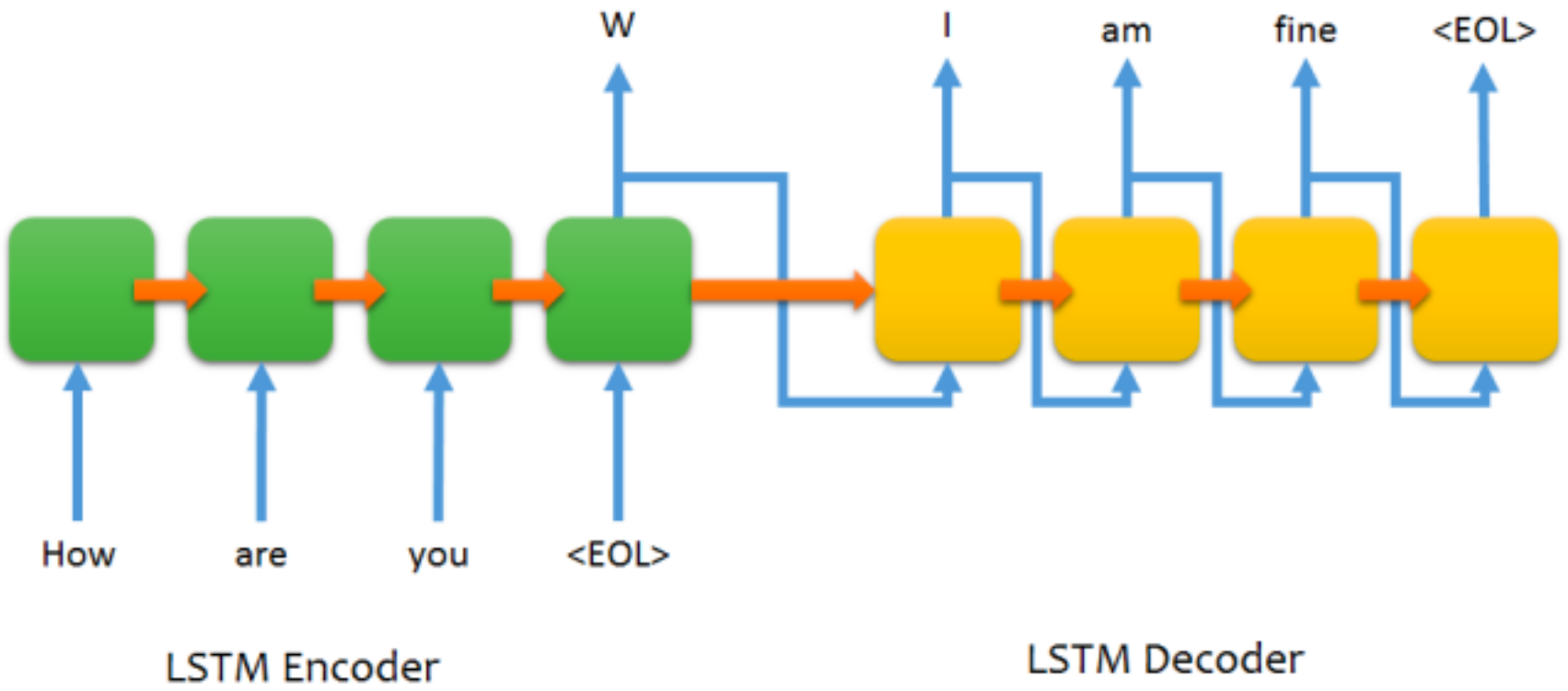


Sequence input and sequence output (e.g. Machine Translation: an RNN reads a sentence in English and then outputs a sentence in French)

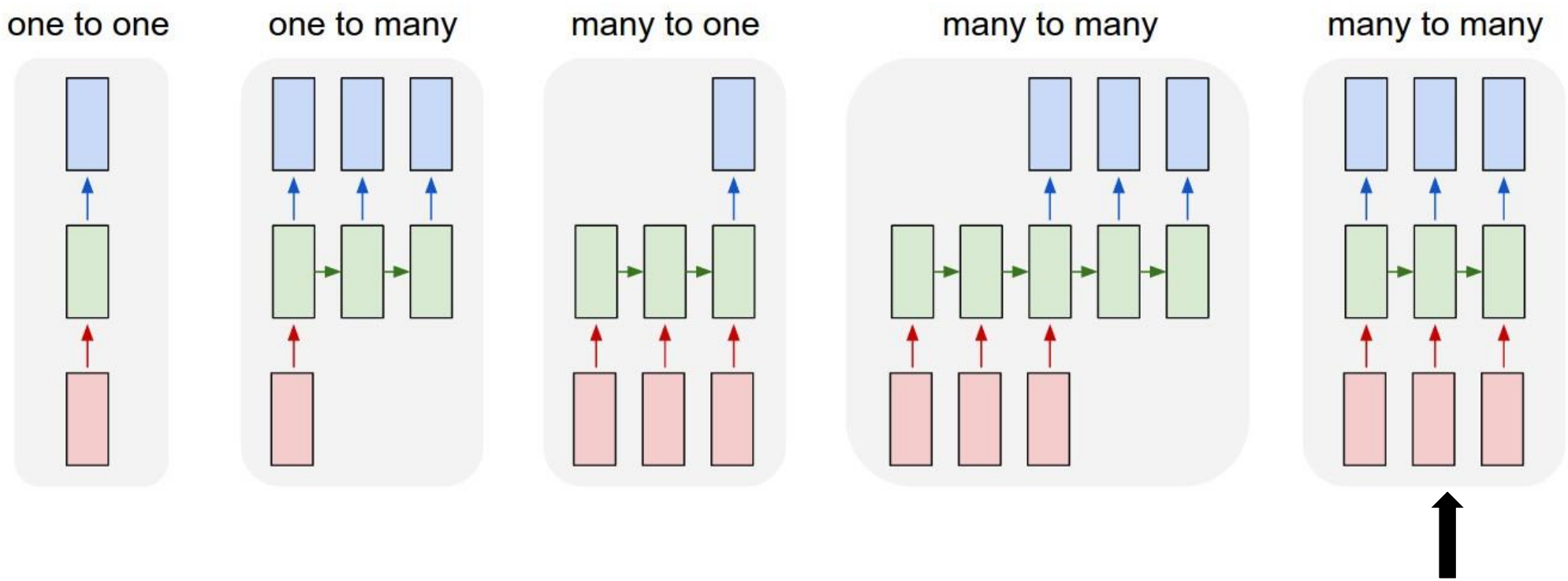
Neural Machine Translation



Sequence to Sequence chat model

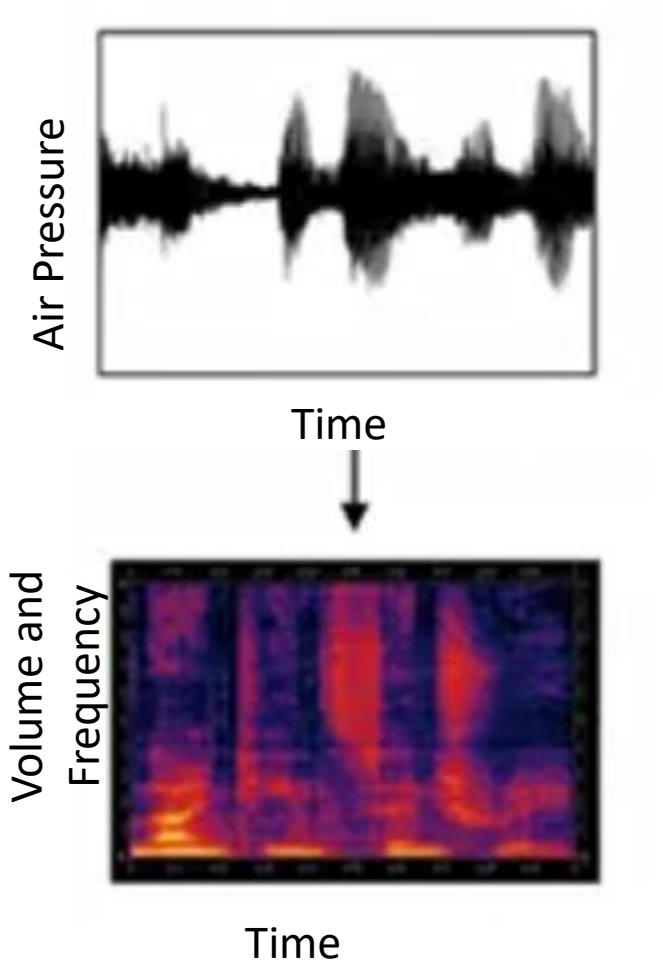


RNN Variants



Synced sequence input and output (e.g. video classification where we wish to label each frame of the video).

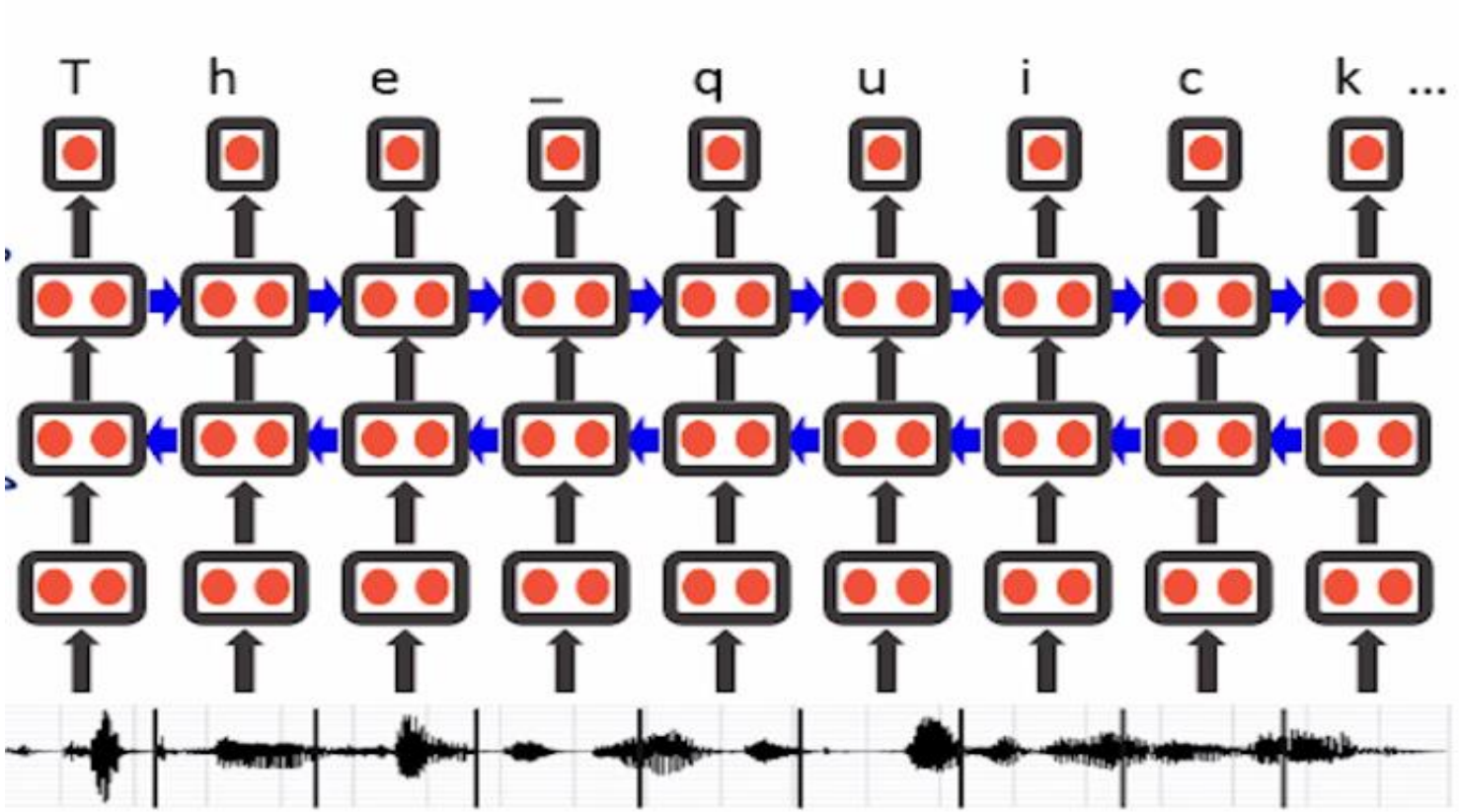
Speech Recognition



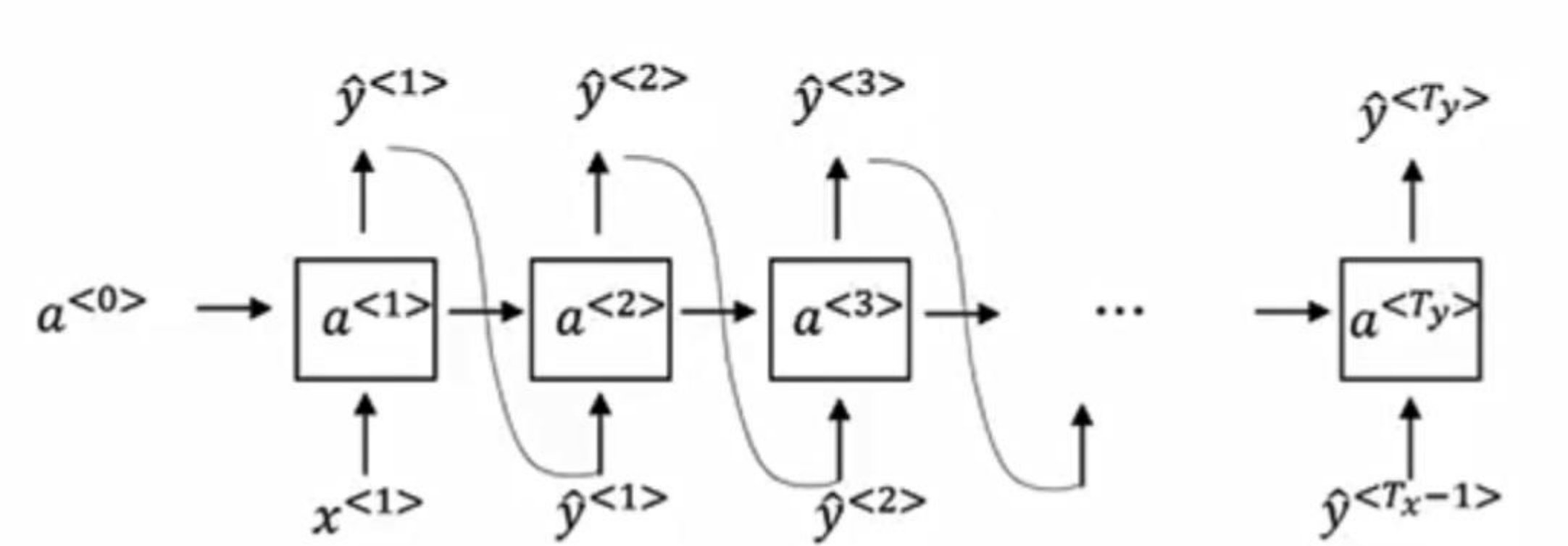
Previously we used to have Hand-engineered features consisting of different phonemes

Thousands of hours of speech/audio data is used to train the data depending upon the application

Baidu’s speech recognition using RNN



Word level and character level language model



Sampling a novel sequence

Once you have a trained model on a corpus you can also have a RNN that can sample new sequences for you.

In that case you initialize with a zero and your first output gives a probability in terms of softmax function of the size of the no of categories equal to the size of your corpus.

You Choose a random word as the first output and then that word acts as the input for the second input and so on.

If you get a <UNK> then you can reject that token and continue with the next guess. It can go on until you get a <EOS> token.

Featurized Representation: Word Embeddings

	Man 5391	Woman 9853	King 4914	Queen 7157	Apple 456	Orange 6257
Gender	1	1	0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0
Age	0.03	0.02	0.7	0.69	0.03	0.02
Food	0.04	0.01	0.02	0.01	0.95	0.97
Size						
Cost						
Alive						

If we have 300 such properties and 10000 Words then it will be a 300x10000 Matrix and is denoted as Embedding Matrix (E) and O_{man} is One hot Vector for Man Word and e_{man} can be embedding vector for Man word.

I Want A glass of orange __Juice
I want a glass of Apple

Transfer Learning and Word Embeddings

- Learn Word Embeddings from Large Text Corpus (1-100 Billion Words)
- Pre-trained embeddings are available online
- Transfer Embedding to a new task with smaller training set
- Continue to finetune word embeddings with new data

Analogies using Word Vectors

As Man-> Woman King->?

As INR->India Dollar->?

As Delhi->India Kathmandu->?

As Tall->taller Big->?

As Man->Woman Boy->?

……

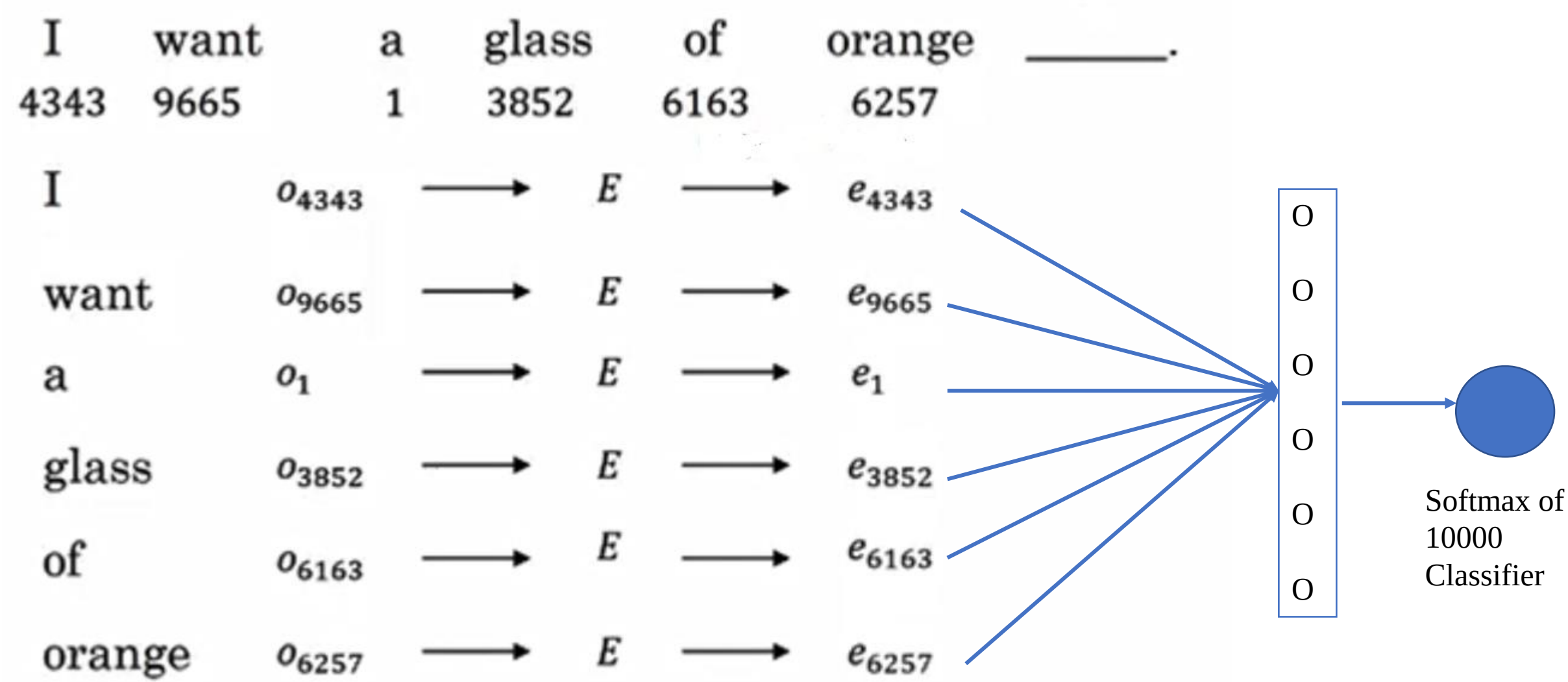
$e_{man}-e_{woman} \approx e_{king}-e_{w}$

Find a word w : Maximize similarity(e_w , $e_{king}-e_{man}+e_{woman}$)

Cosine similarity

Sim(u,v)= $u^T v$ / ||u|| ||v||

Neural Language Model



Removing Biases in NLP

- It related to Gender and ethnicity biases and we need to be very careful about this
- Man: Computer_Programmer as Women Homemaker
- Father: Doctor Mother: Nurse

- Biases will be picked from the text it has been trained upon
- First step is to identify Bias Direction e.g. male to female
- Next is to Neutralize the bias for all the non-definitional word for example Father, Mother, He, She are definitional word for Gender and should not get changes due to this. However, Non-definitional word like soldier, doctor, Manager, Programmer etc should be neutralized for bias
- Last step is to equalize pairs like niece, nephew; grandmother, grandfather and they should be equidistant from words like babysitter etc.

Public Notice regarding Use of Images/Information

This document contains images obtained by routine Google Images searches. Some of these images may perhaps be under copyright. They are included here for educational and noncommercial purposes and are considered to be covered by the doctrine of Fair Use. In any event they are easily available from Google Images.

It's not feasible to give full scholarly credit to the creators of the images/information. We hope they can be satisfied with the positive role they are playing in the educational process.


Thank
you!