

COMP5046

Natural Language Processing

Lecture 13: Content Review and Exam Guide

Semester 1, 2020

School of Computer Science

The University of Sydney, Australia



THE UNIVERSITY OF
SYDNEY

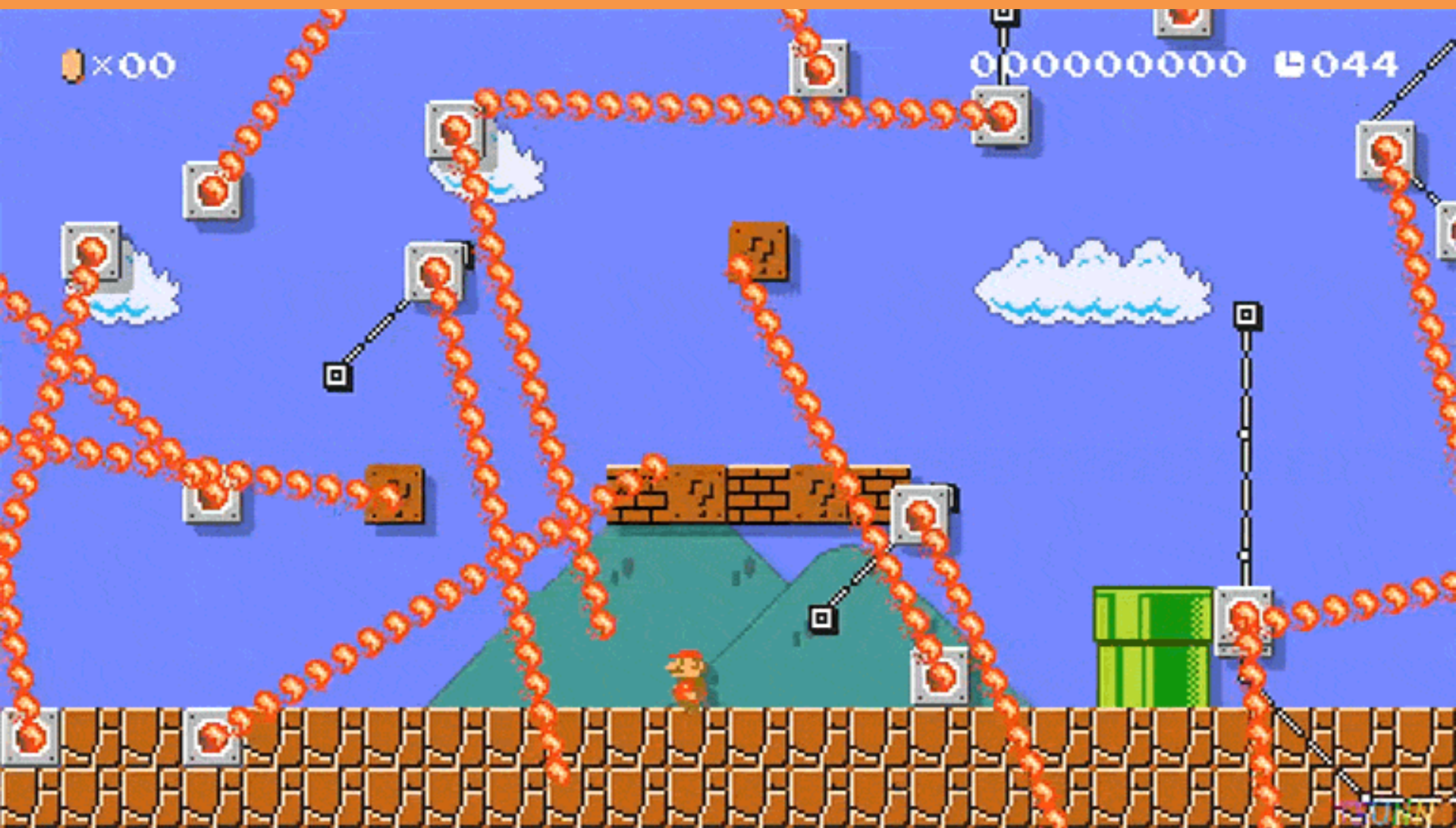
Dr Caren Han

Caren.Han@sydney.edu.au

Lecture 13: Content Review and Exam Guide

1. Overview
2. Learning Content Summary
3. Final Exam Guide

Your Journey...!



You are a survivor!!

A close-up shot of a man with a grey beard and short dark hair, looking slightly to his right with a surprised expression. He is wearing a light blue button-down shirt. The background is dark and out of focus, with a warm, orange light source visible on the right side.

I can't believe I did that.

Survey (5 mins)

Have your say! 😊

If you are reading this outside of the lecture, please do this now!

- *Unit of Study Evaluation survey at:*

<https://student-surveys.sydney.edu.au/students/>

- *Your answers are completely anonymous*
- *Please write specific comments*
- *I personally read every single comment*

What we learned in this course!

Week 1: Introduction to Natural Language Processing (NLP)

Week 2: Word Embeddings (Word Vector for Meaning)

Week 3: Word Classification with Machine Learning I

Week 4: Word Classification with Machine Learning II

NLP and
Machine
Learning

Week 5: Language Fundamental

Week 6: Part of Speech Tagging

Week 7: Dependency Parsing

Week 8: Language Model

NLP
Techniques

Week 9: Information Extraction: Named Entity Recognition

Week 10: Advanced NLP: Attention and Reading Comprehension

Week 11: Advanced NLP: Transformer and Machine Translation

Week 12: Advanced NLP: Graph Neural Network with NLP

Advanced
Topic

Week 13: Future of NLP and Exam Review

What we learned in this course!

Week 1: Introduction to Natural Language Processing (NLP)

Week 2: Word Embeddings (Word Vector for Meaning)

Week 3: Word Classification with Machine Learning I

Week 4: Word Classification with Machine Learning II

**NLP and
Machine
Learning**

Week 5: Language Fundamental

Week 6: Part of Speech Tagging

Week 7: Dependency Parsing

Week 8: Language Model

**NLP
Techniques**

Week 9: Information Extraction: Named Entity Recognition

Week 10: Advanced NLP: Attention and Reading Comprehension

Week 11: Advanced NLP: Transformer and Machine Translation

Week 12: Advanced NLP: Graph Neural Network with NLP

**Advanced
Topic**

Week 13: Future of NLP and Exam Review

2

Lecture Summary

Lecture 1 Summary

Word Representation with WordNet

- Use e.g. WordNet, a thesaurus containing lists of synonym sets and hypernyms (“is a” relationships).
- Problems with resources like WordNet

Representing words as discrete symbols

- One-hot vector 😊

motel = [0 0 0 0 0 0 0 0 0 0 0 **1** 0 0 0 0]

hotel = [0 0 0 0 0 0 0 0 **1** 0 0 0 0 0 0 0]

In web search, if user searches for “**Seattle motel**”, we would like to match documents containing “**Seattle hotel**”

- Problems with words as discrete symbols

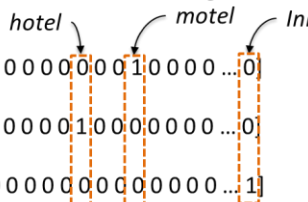
Lecture 1 Summary

Count based Word representation

Important to know: how each techniques works / pros and cons

- **One-hot encoding**

$\text{motel} = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ \dots\ 0]$
 $\text{hotel} = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ \dots\ 0]$
 $\text{Inn} = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \dots\ 1]$



- **Bag of Words**

[1, 1, 1, 1, 2, 2, 2, 1, 1, 3]

a	are	been	day	have	how	nice	see	to	you
1	1	1	1	2	2	2	1	1	3



- **Term Frequency-Inverse Document Frequency**

TF is a count of how many times a word occurs in a given document (= bag of words)

IDF is the number of times a word occurs in a corpus of documents

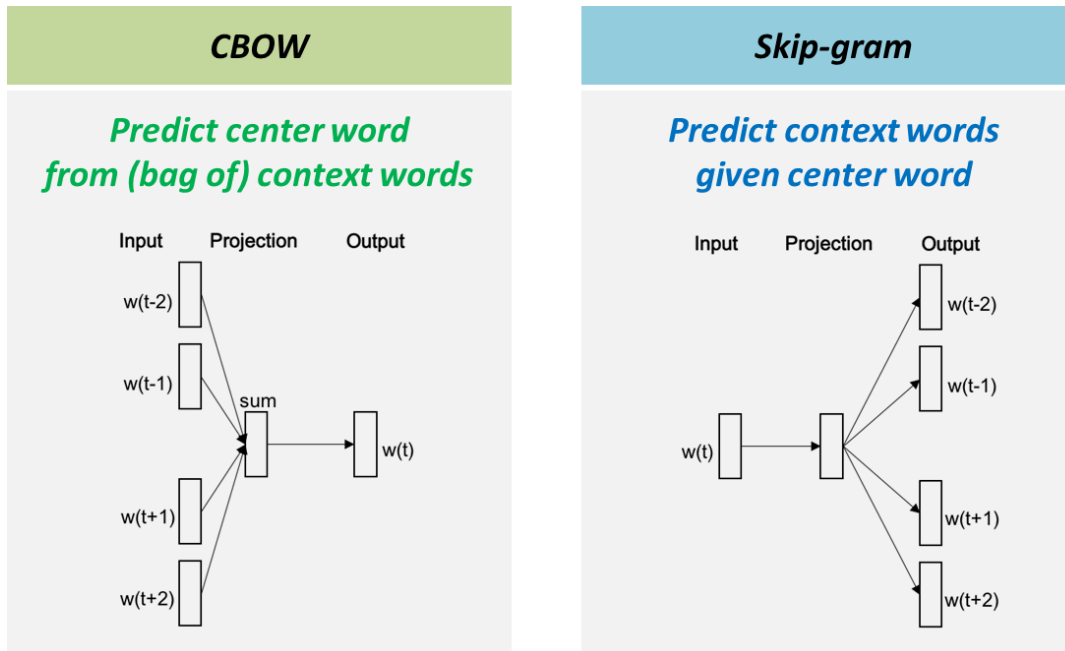
$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

Lecture 2 Summary

Prediction-based Word Representation

Important to know: how each techniques works / pros and cons

- **Word2Vec (CBOW, Skip-Gram)**



Key parameter for training methods: window size, negative samples

Lecture 2 Summary

Prediction-based Word Representation

Important to know: how each techniques works / pros and cons

- ***FastText with n-gram embedding***



- ***Glove***

“Methods like skip-gram may do better on the analogy task, but they poorly utilize the statistics of the corpus since they train on separate local context windows instead of on global co-occurrence counts.”

Lecture 3 Summary

Word Embedding Evaluation

- Intrinsic Evaluation
- Extrinsic Evaluation

Type	How to work / Benefit
Intrinsic	Evaluation on a specific/intermediate subtask
	<ul style="list-style-type: none">• Fast to compute• Helps to understand that system• Not clear if really helpful unless correlation to real task is established
Extrinsic	Evaluation on a real task
	<ul style="list-style-type: none">• Can take a long time to compute accuracy• Unclear if the subsystem is the problem or its interaction or other subsystems

Lecture 3 Summary

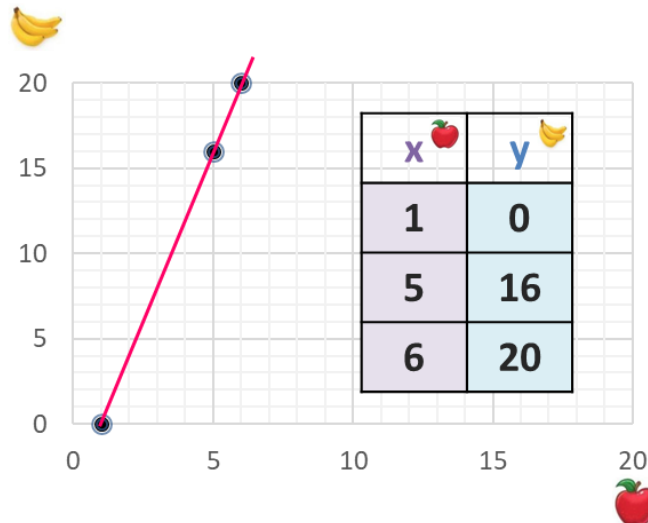
Deep Learning for Natural Language Processing

- Perceptron and Neural Network (NN)
- Multilayer Perceptron

Input: x =number of apple given by Lisa

Output: y =number of banana received by Lisa

Parameters: Need to be estimated



Model

$$y = \overset{\text{weight}}{W} x + \overset{\text{bias}}{b}$$

How can we find the parameters, w and b ?

Lecture 3 Summary

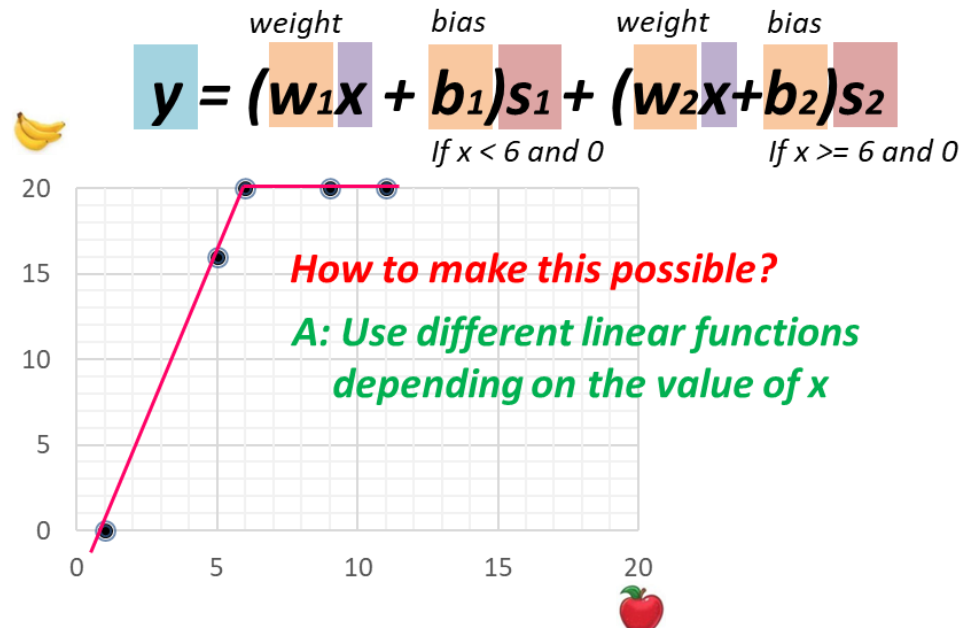
Deep Learning for Natural Language Processing

- Perceptron and Neural Network (NN)
- Multilayer Perceptron

Nonlinear Neural Network

Data

x 🍎	y 🍌
1	0
5	16
6	20
9	20
11	20

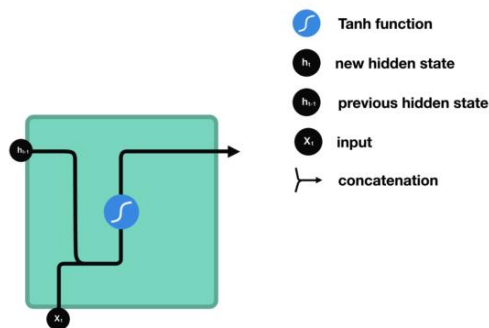


Lecture 4 Summary

Seq2Seq Learning

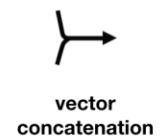
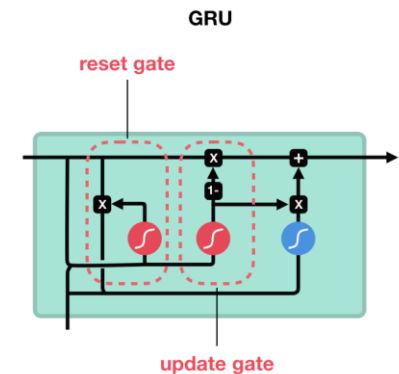
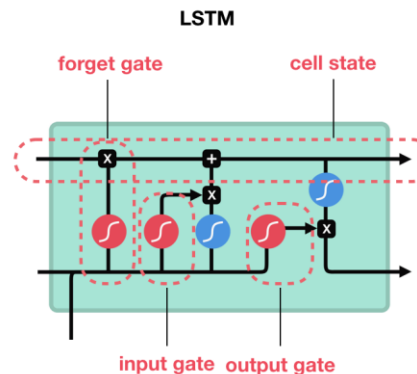
Seq2Seq Deep Learning

1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)



$$h_t = f_W(h_{t-1}, x_t)$$

New hidden state h_t is calculated as a function f_W with parameters W applied to the previous state h_{t-1} and input x_t .

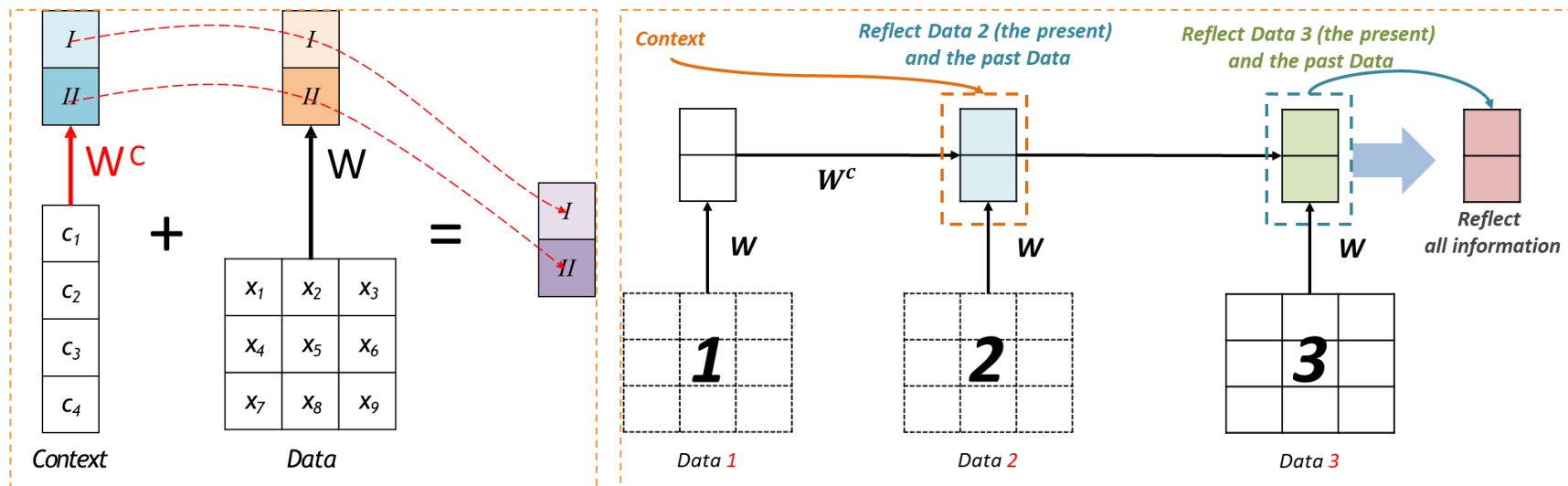


Lecture 4 Summary

Seq2Seq Learning

Seq2Seq Deep Learning

1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)

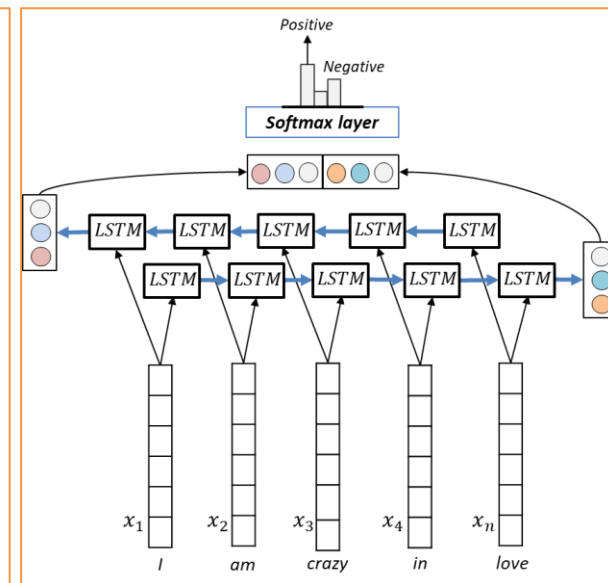
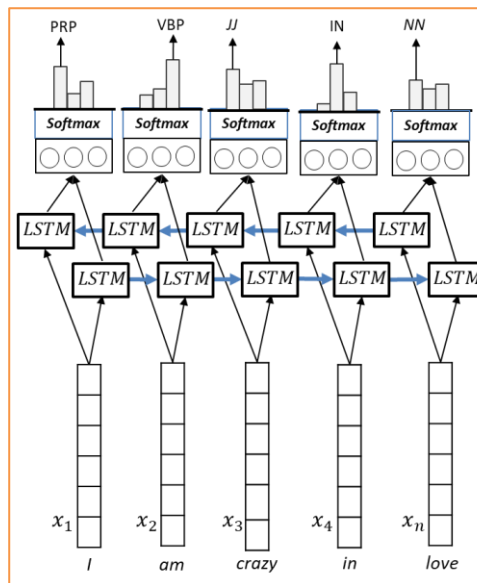
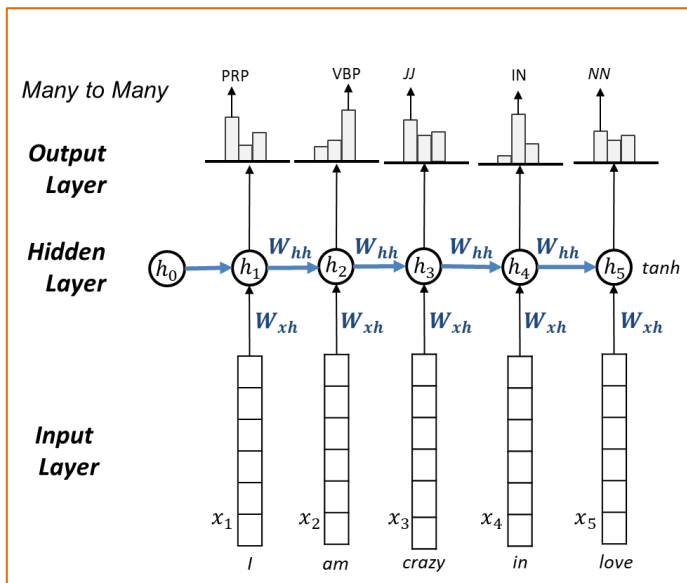


Lecture 4 Summary

Seq2Seq Learning

Seq2Seq Deep Learning

1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)



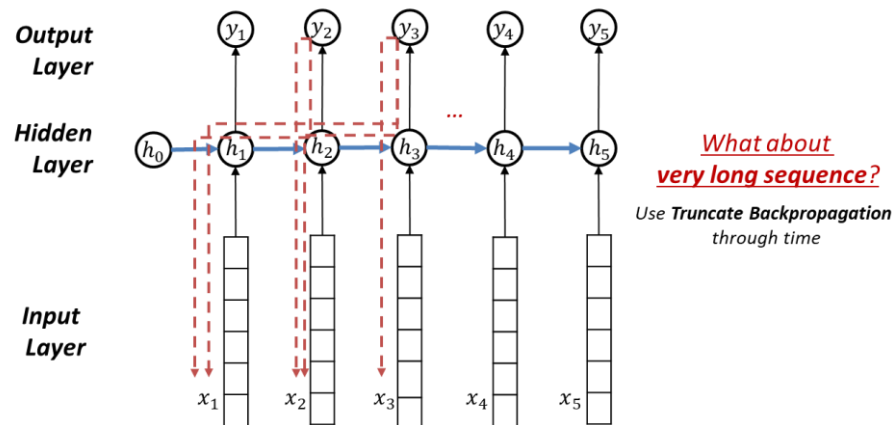
Lecture 4 Summary

Seq2Seq Learning

Seq2Seq Deep Learning

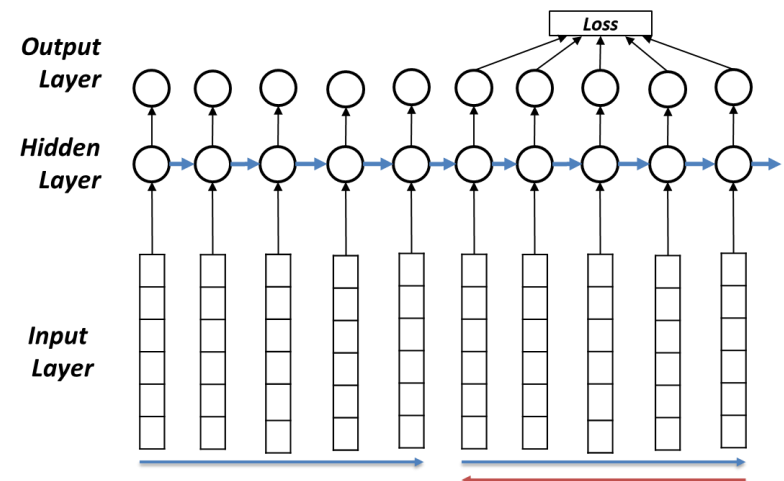
1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)

Backpropagation through time



- Similar as **standard backpropagation** on unrolled network
- Similar as **training very deep networks** with tied parameters

Truncated Backpropagation through time



Carry hidden states forward in time forever, but only backpropagate for some smaller number of steps

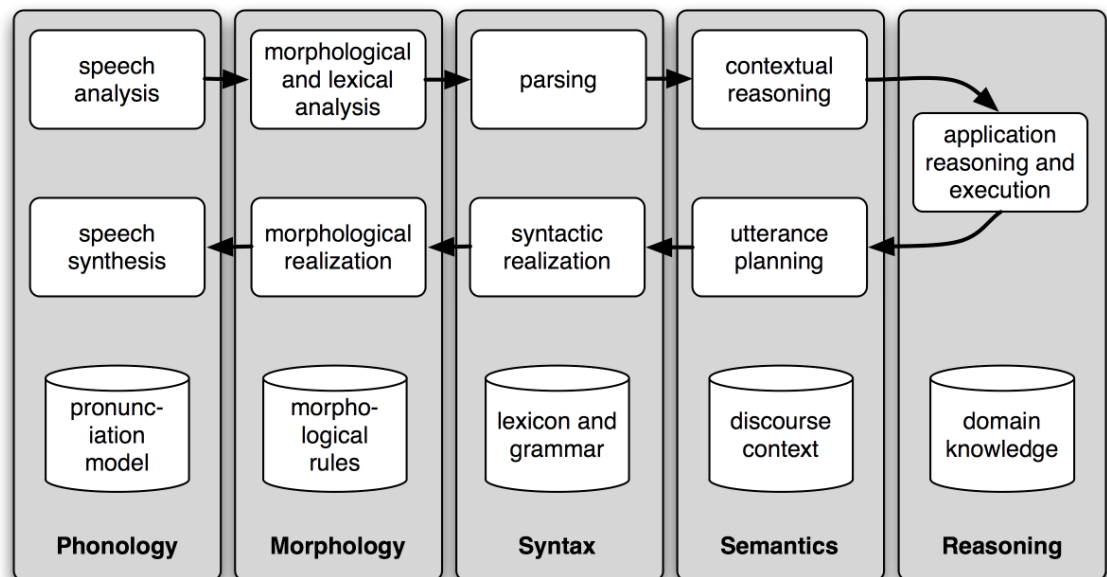
Lecture 5 Summary

Language Fundamental

1. Phonology, Morphology, Syntax, Semantics, Pragmatics

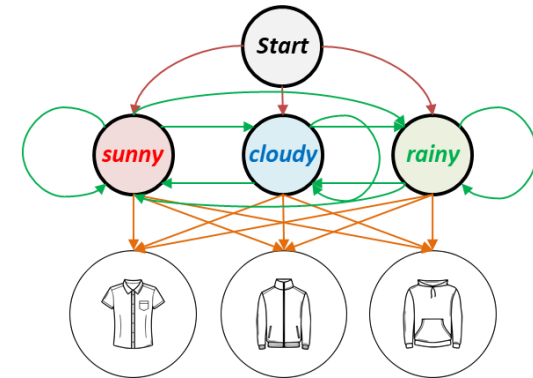
Text Preprocessing

1. Tokenization
2. Cleaning and Normalisation
3. Stemming and Lemmatisation
4. Stopword
5. Regular Expression



Lecture 6 Summary

1. Part-of-Speech Tagging
2. Baseline Approaches
 1. Rule-based Model
 2. Look-up Table Model
 3. N-Gram Model
3. Probabilistic Approaches
 1. Hidden Markov Model
 2. Conditional Random Field
4. Deep Learning Approaches



Predicting the **weather (hidden variable)** based on the type of **clothes that someone wears (observed)**

Initial

Rainy	0.6
Cloudy	0.3
Sunny	0.1

Transitions

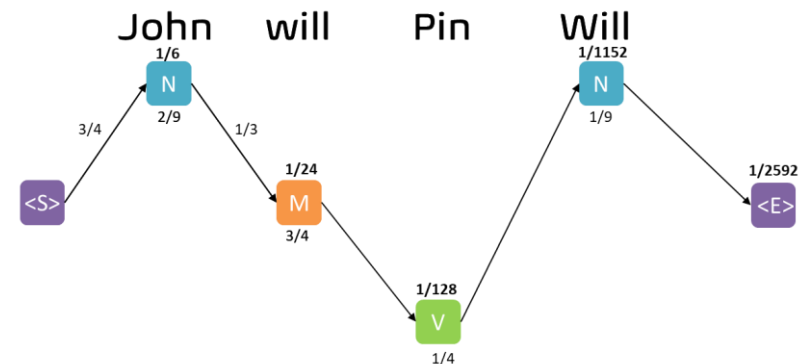
		Tomorrow		
		Rainy	Cloudy	Sunny
Today	Rainy	0.6	0.3	0.1
	Cloudy	0.4	0.3	0.3
	Sunny	0.1	0.4	0.5

Emissions

	Shirts	Jacket	Hoodies
Rainy	0.8	0.19	0.01
Cloudy	0.5	0.4	0.1
Sunny	0.01	0.2	0.79

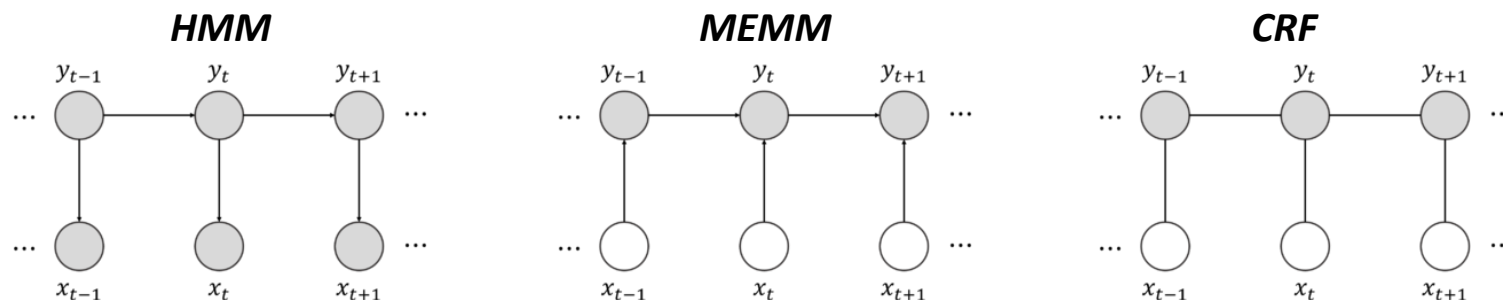
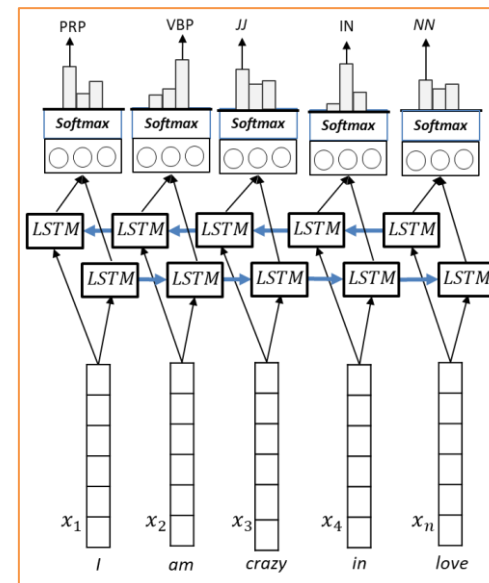
	N	V	M
Emma	4/9	0	0
John	2/9	0	0
Will	1/9	0	3/4
Pin	2/9	1/4	0
Can	0	0	1/4
Meet	0	2/4	0
Pat	0	1/4	0

	N	V	M	<E>
<S>	3/4	0	1/4	0
N	1/9	1/9	3/9	4/9
V	4/4	0	0	0
M	1/4	3/4	0	0



Lecture 6 Summary

1. Part-of-Speech Tagging
2. Baseline Approaches
 1. Rule-based Model
 2. Look-up Table Model
 3. N-Gram Model
3. Probabilistic Approaches
 1. Hidden Markov Model
 2. Conditional Random Field
4. Deep Learning Approaches



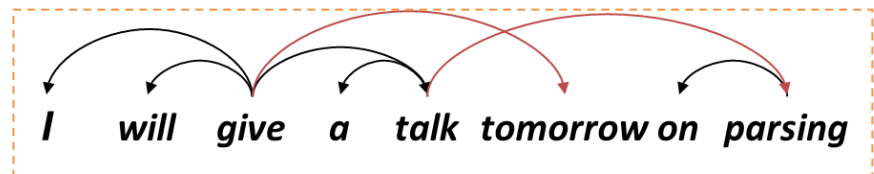
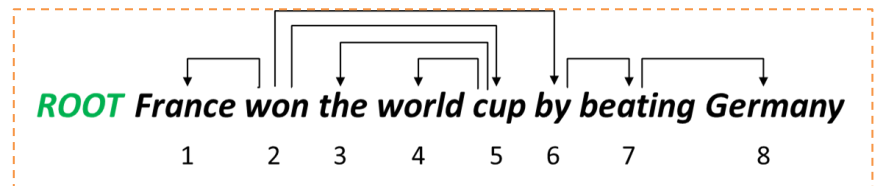
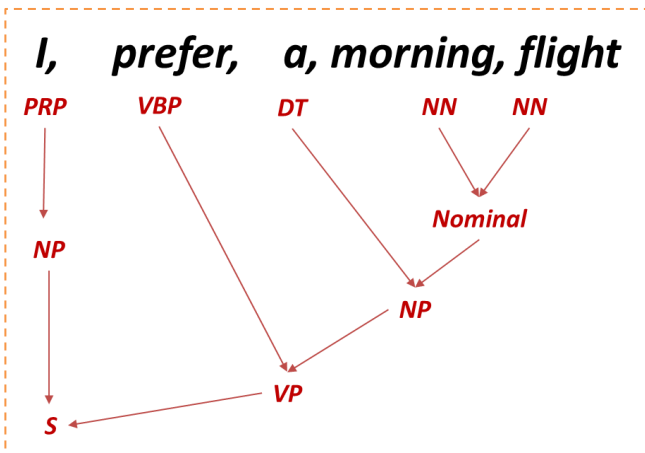
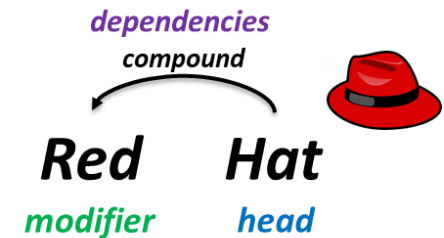
Lecture 7 Summary

Linguistic Structure

Dependency Structure

Dependency Parsing Algorithms

1. Transition-based Dependency Parsing
2. Deep Learning-based Dependency Parsing



Lecture 7 Summary

Linguistic Structure

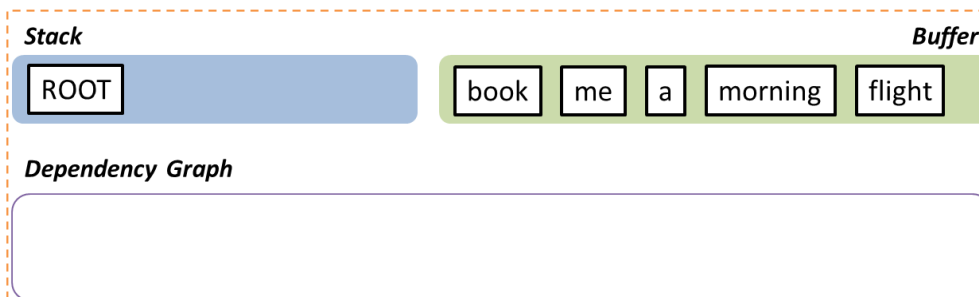
Dependency Structure

Dependency Parsing Algorithms

1. Transition-based Dependency Parsing
2. Deep Learning-based Dependency Parsing

***Graph-based
Dependency Parsing***

***Transaction-based
Dependency Parsing***



How to choose next action?

****Unlabeled Attachment Score***
****Labeled Attachment Score***

Lecture 8 Summary

Language Model

Traditional Language Model

Neural Language Model

Natural Language Generation

Other NLG Approaches

Language Model and NLG Evaluation

$P(\text{An, adorable, little, boy, is, spreading, smiles})$

$= P(\text{An}) \times P(\text{adorable}|\text{An}) \times P(\text{little}|\text{An adorable}) \times P(\text{boy}|\text{An adorable little}) \times P(\text{is}|\text{An adorable little boy}) \times P(\text{spreading}|\text{An adorable little boy is}) \times P(\text{smiles}|\text{An adorable little boy is spreading})$

Trained Corpus

An adorable little boy is

..... An adorable little boy is

An adorable little boy laughed

- ***Count-based***
- ***N-gram Language Models***

Lecture 8 Summary

Language Model

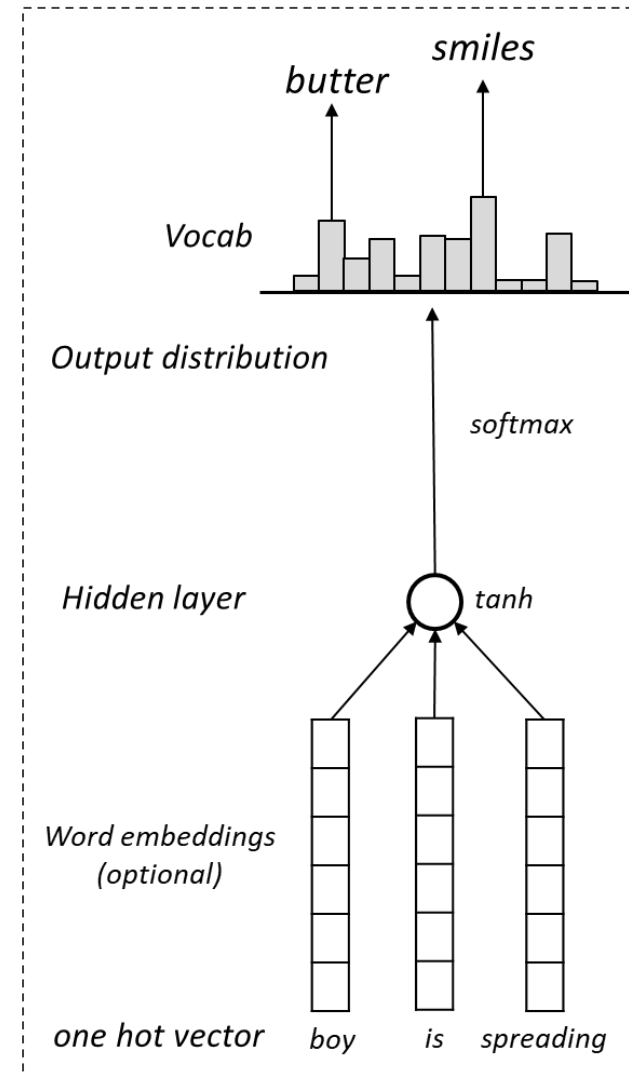
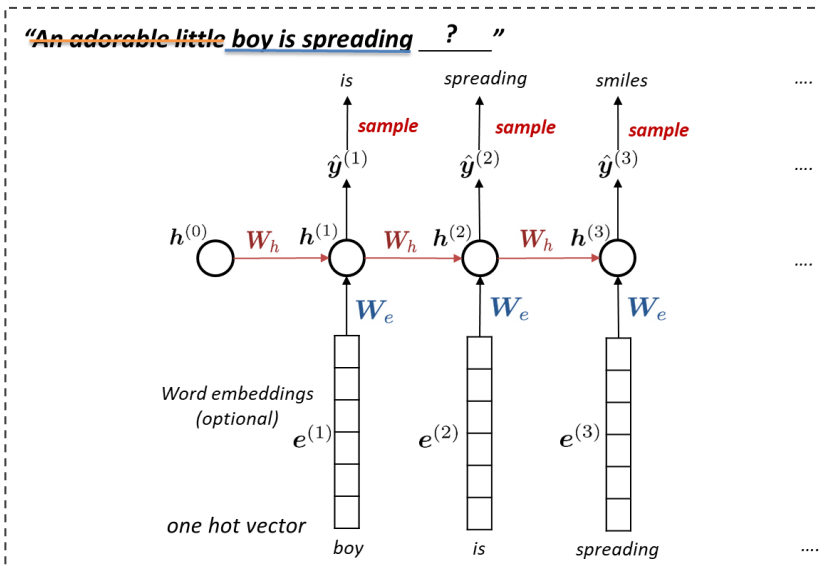
Traditional Language Model

Neural Language Model

Natural Language Generation

Other NLG Approaches

Language Model and NLG Evaluation



Lecture 8 Summary

Language Model

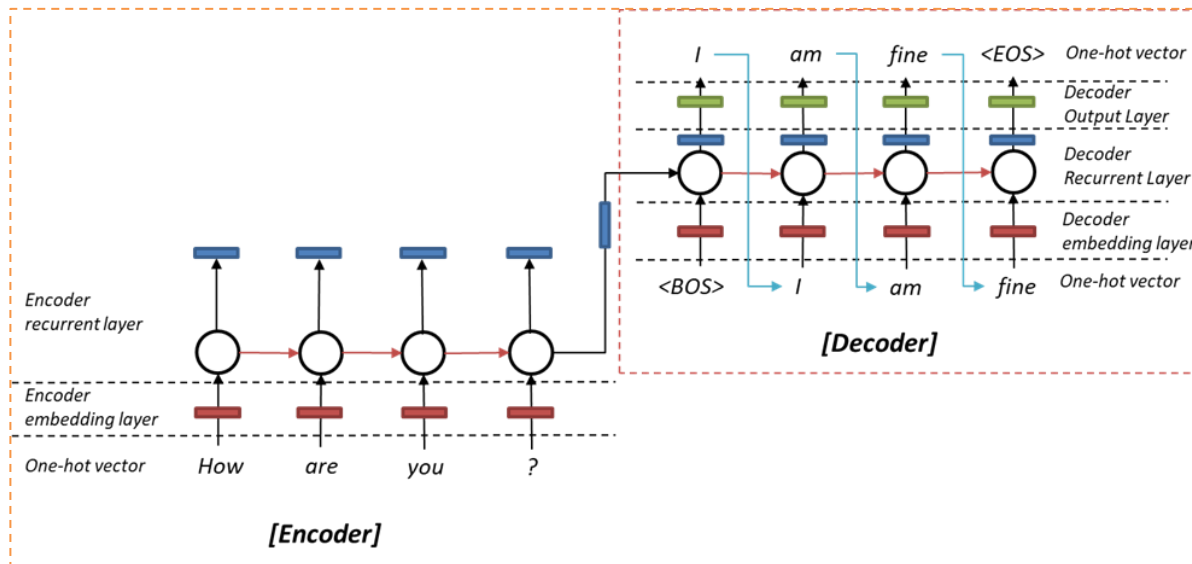
Traditional Language Model

Neural Language Model

Natural Language Generation

Other NLG Approaches

Language Model and NLG Evaluation



Decoding Algorithm

1: Greedy Decoding

2: Beam Search

Lecture 8 Summary

Language Model

Traditional Language Model

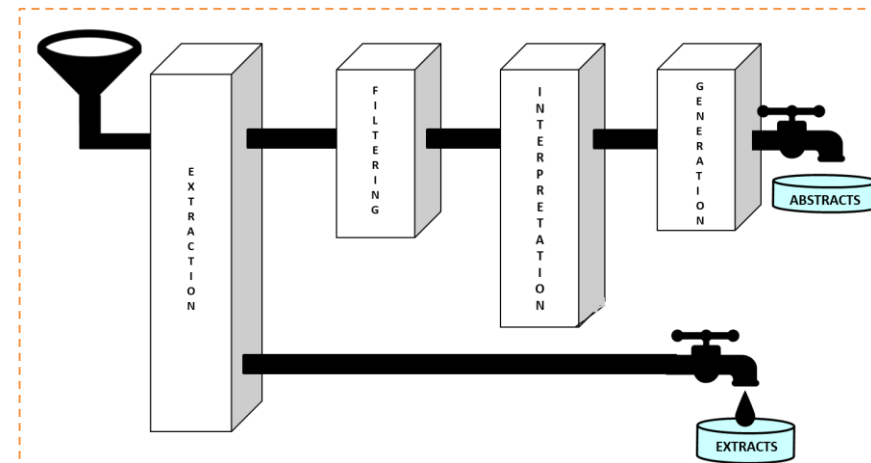
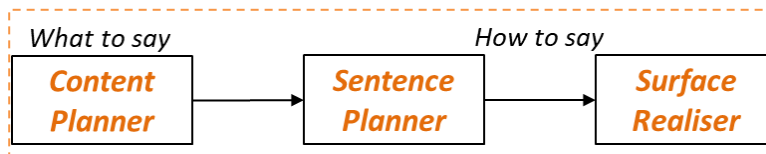
Neural Language Model

Natural Language Generation

Other NLG Approaches

Language Model and NLG Evaluation

“Flights from **ORIGIN** to **DEST** on **DEPT_DATE**
DEPT_TIME. Just one moment please”



Lecture 8 Summary

Language Model

Traditional Language Model

Neural Language Model

Natural Language Generation

Other NLG Approaches

Language Model and NLG Evaluation

Perplexity

$$\text{perplexity} = \prod_{t=1}^T \left(\frac{1}{P_{\text{LM}}(\mathbf{x}^{(t+1)} | \mathbf{x}^{(t)}, \dots, \mathbf{x}^{(1)})} \right)^{1/T}$$

So, Lower Perplexity is better!

Lecture 9 Summary

Information Extraction

Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation

Rule-based NER

1. *Determining which person holds what office in what organization*
2. *Determining where an organization is located*

Unigram	Mr.	Scott	Morrison	flew	to	Beijing
Lowercase unigram	mr.	scott	morrison	flew	to	beijing
POS tag	nnp	nnp	nnp	vbd	to	nnp
length	3	5	4	4	2	7
In first-name gazetteer	no	yes	no	no	no	no
In location gazetteer	no	no	no	no	no	yes
3-letter suffix	Mr.	ott	son	lew	-	ing
2-letter suffix	r.	tt	on	ew	to	ng
1-letter suffix	.	t	n	w	o	g
Tag predictions	O	B-per	I-per	O	O	B-loc

Lecture 9 Summary

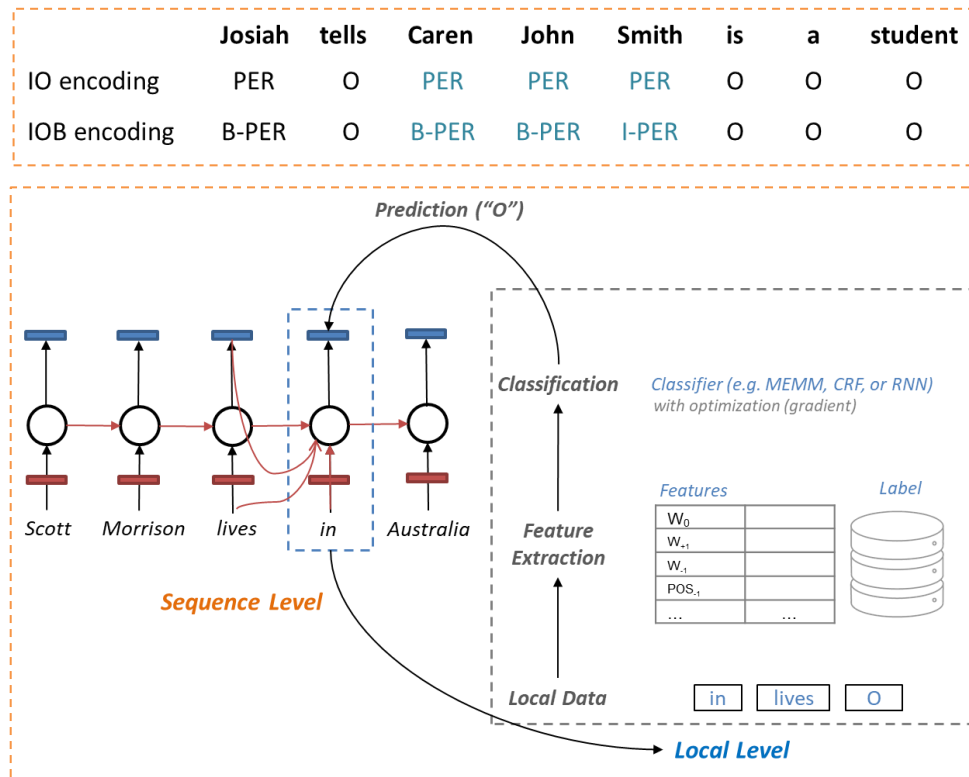
Information Extraction

Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



Lecture 9 Summary

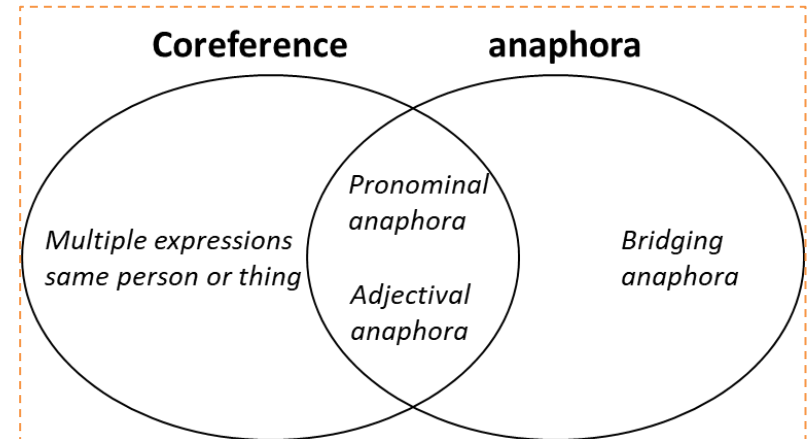
Information Extraction

Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



Ivanka

Donald

he

her

she

Ivanka was happy that **Donald** said **he** considered nominating **her** because **she** is very good with numbers

Gold cluster 1

Gold cluster 2

Lecture 9 Summary

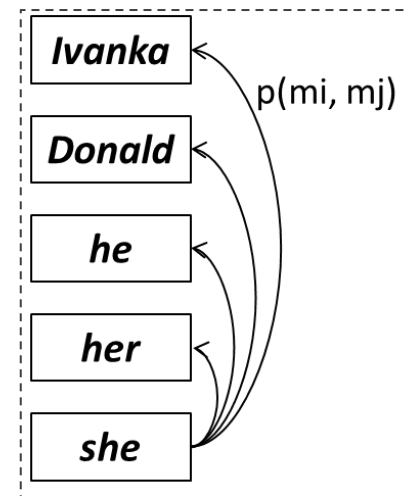
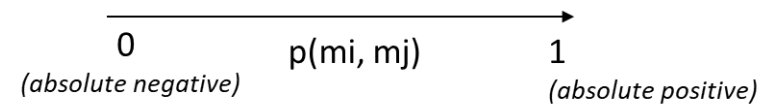
Information Extraction

Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



Lecture 9 Summary

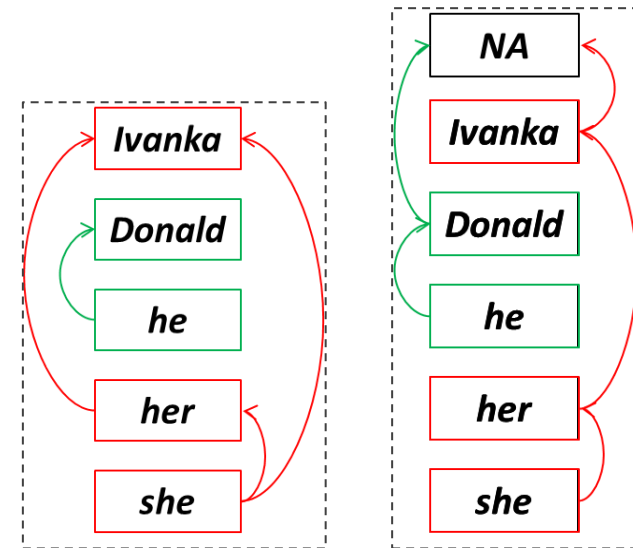
Information Extraction

Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



$$J = - \sum_{i=2}^N \sum_{j=1}^i y_{ij} \log p(m_j, m_i)$$

Iterate through mentions Iterate through candidate antecedents (previously occurring mentions)

Coreferent mentions pairs should get high probability, others should get low probability

Lecture 9 Summary

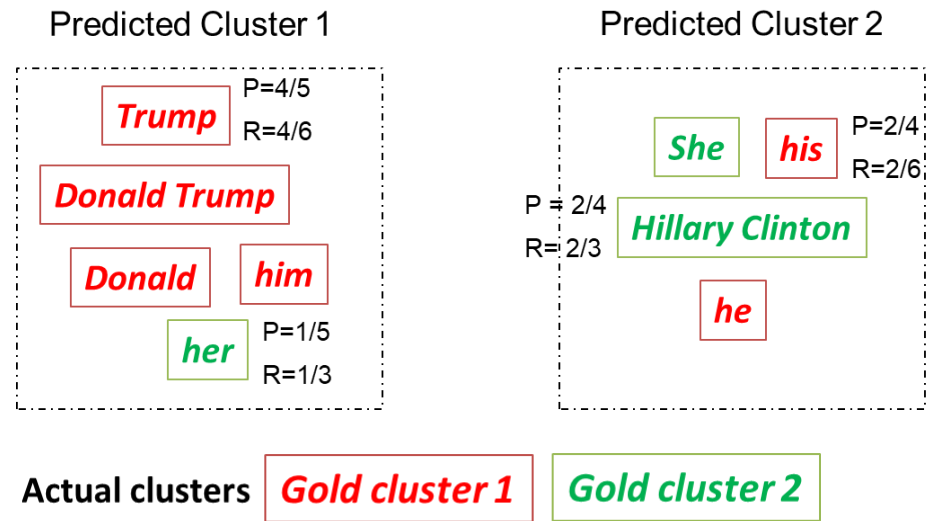
Information Extraction

Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



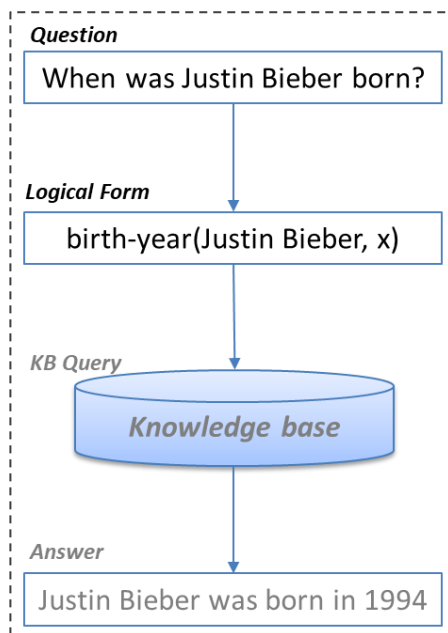
Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

IR-based Question Answering and Reading Comprehension

Additional: Visual Question Answering



Question	Logical Form
When was Justin Bieber born?	<code>birth-year(Justin Bieber, x)</code>
What is the largest state?	<code>argmax(λx.state(x), λx.size(x))</code>

Subject	Predicate (relation)	Object
Justin Bieber	birth-year	1994
Frédéric Chopin	birth-year	1810
...

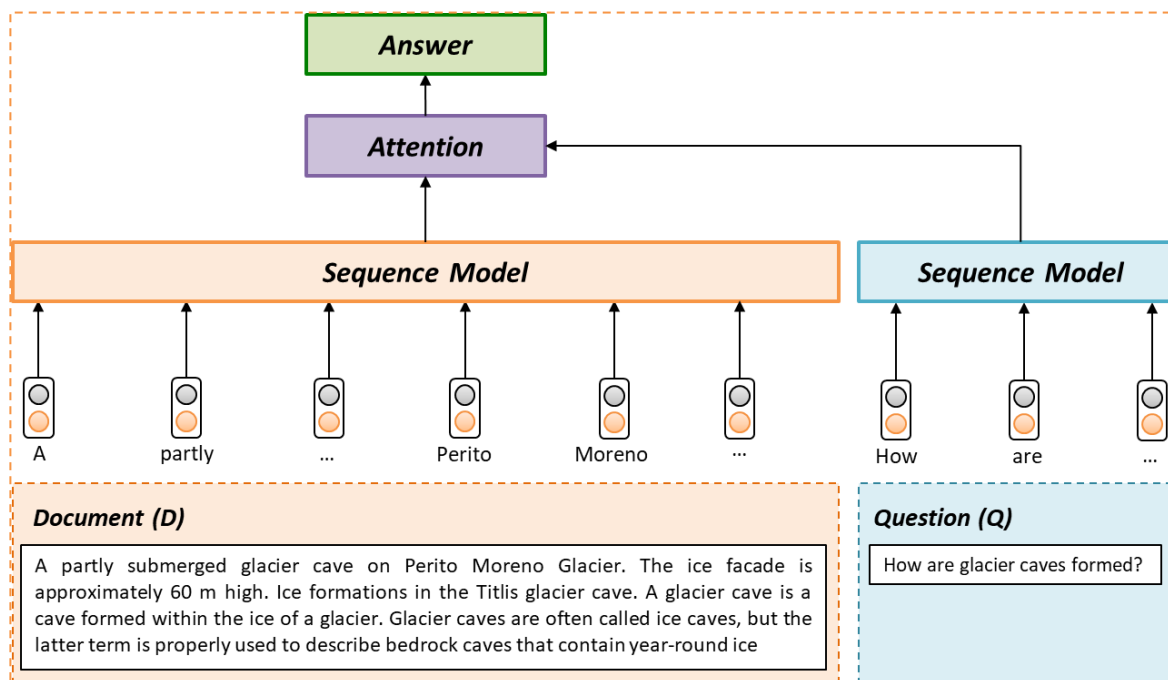
Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

IR-based Question Answering and Reading Comprehension

Additional: Visual Question Answering



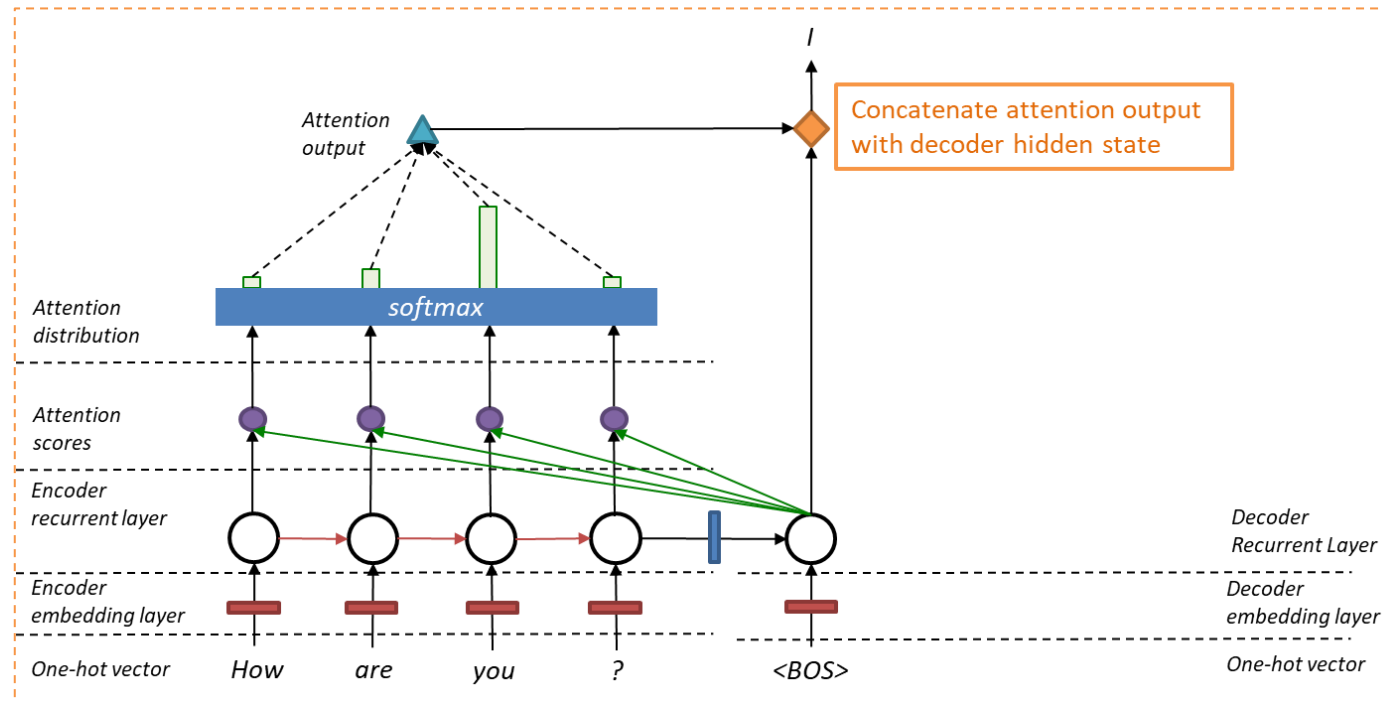
Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

IR-based Question Answering and Reading Comprehension

Additional: Visual Question Answering



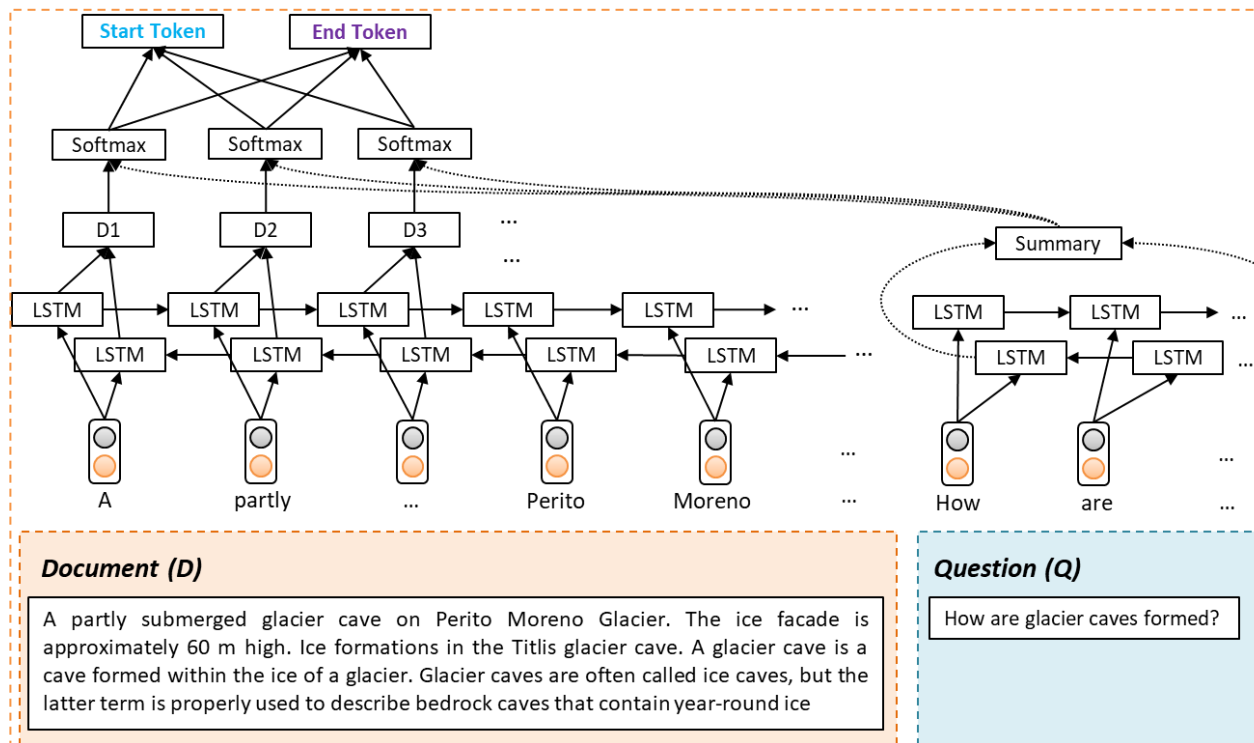
Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

IR-based Question Answering and Reading Comprehension

Additional: Visual Question Answering



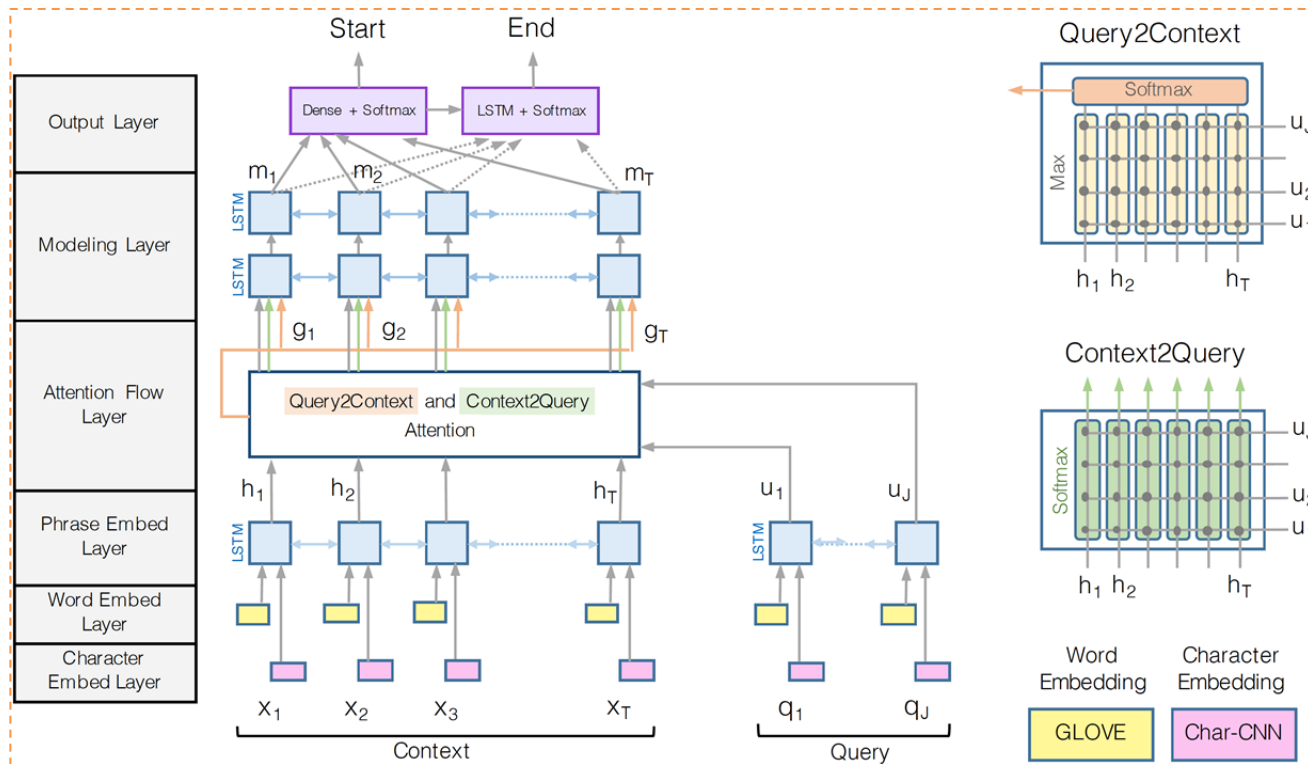
Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

IR-based Question Answering and Reading Comprehension

Additional: Visual Question Answering

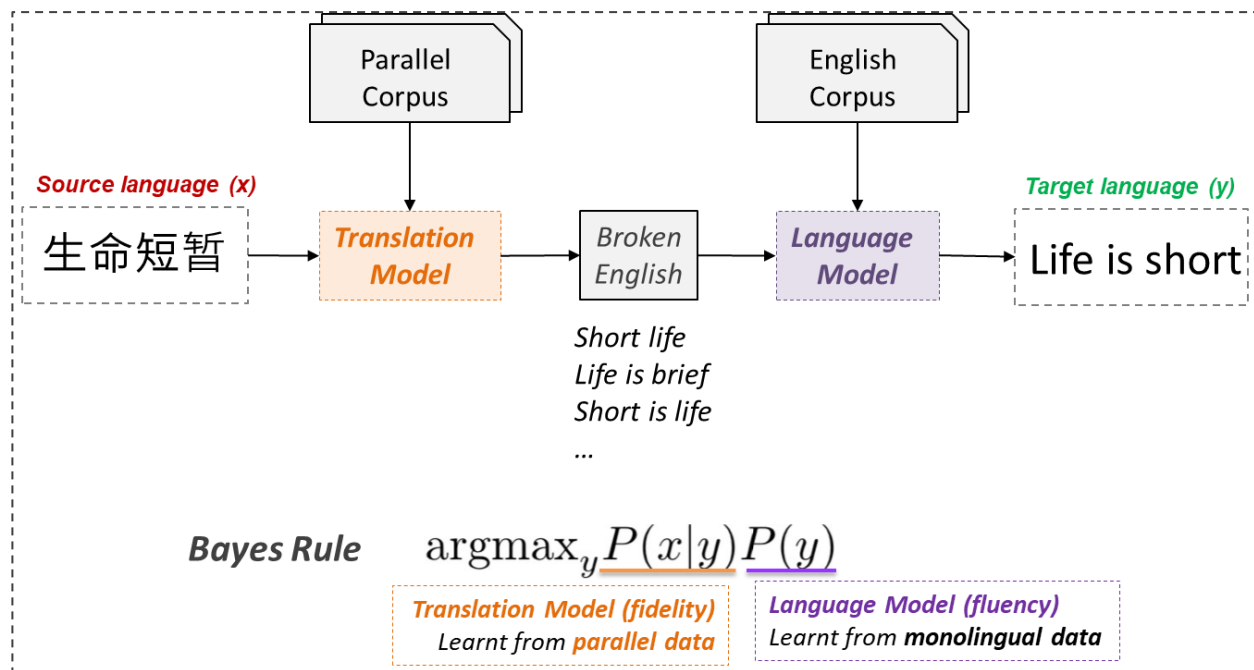


Lecture 11 Summary

Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT

The Rise of the Pre-trained Model

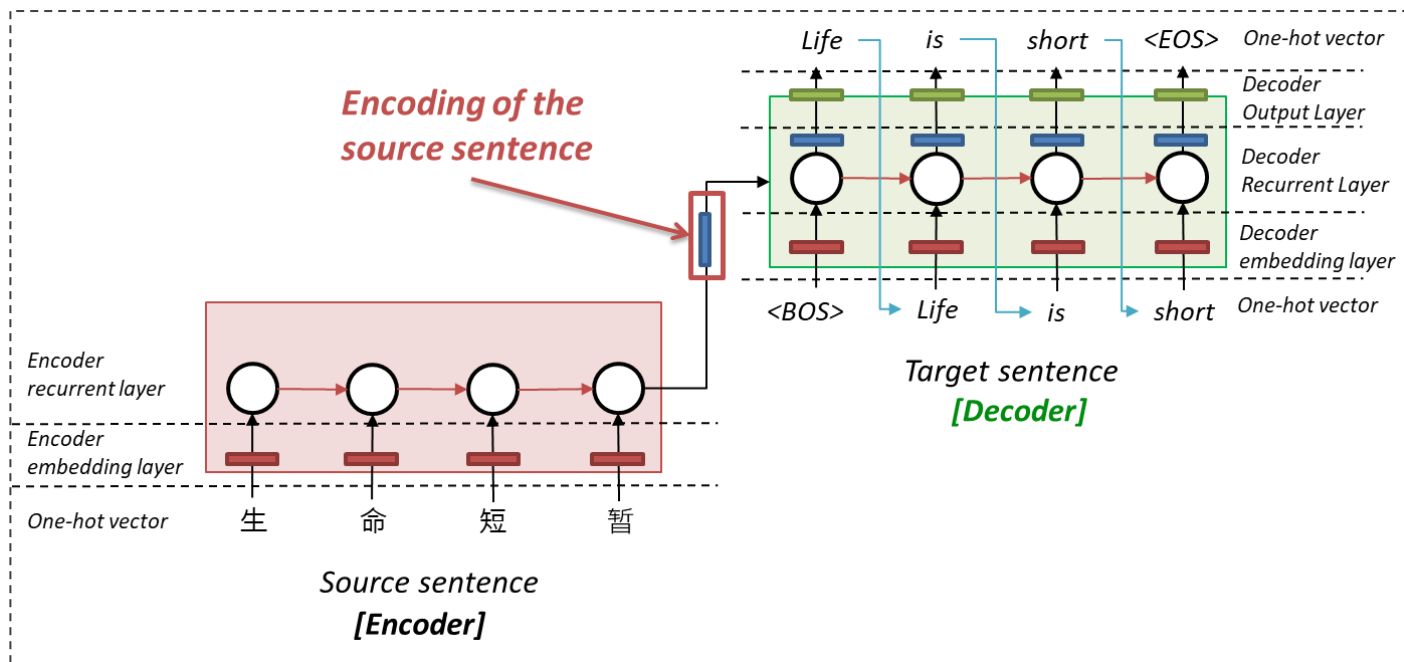


Lecture 11 Summary

Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT

The Rise of the Pre-trained Model

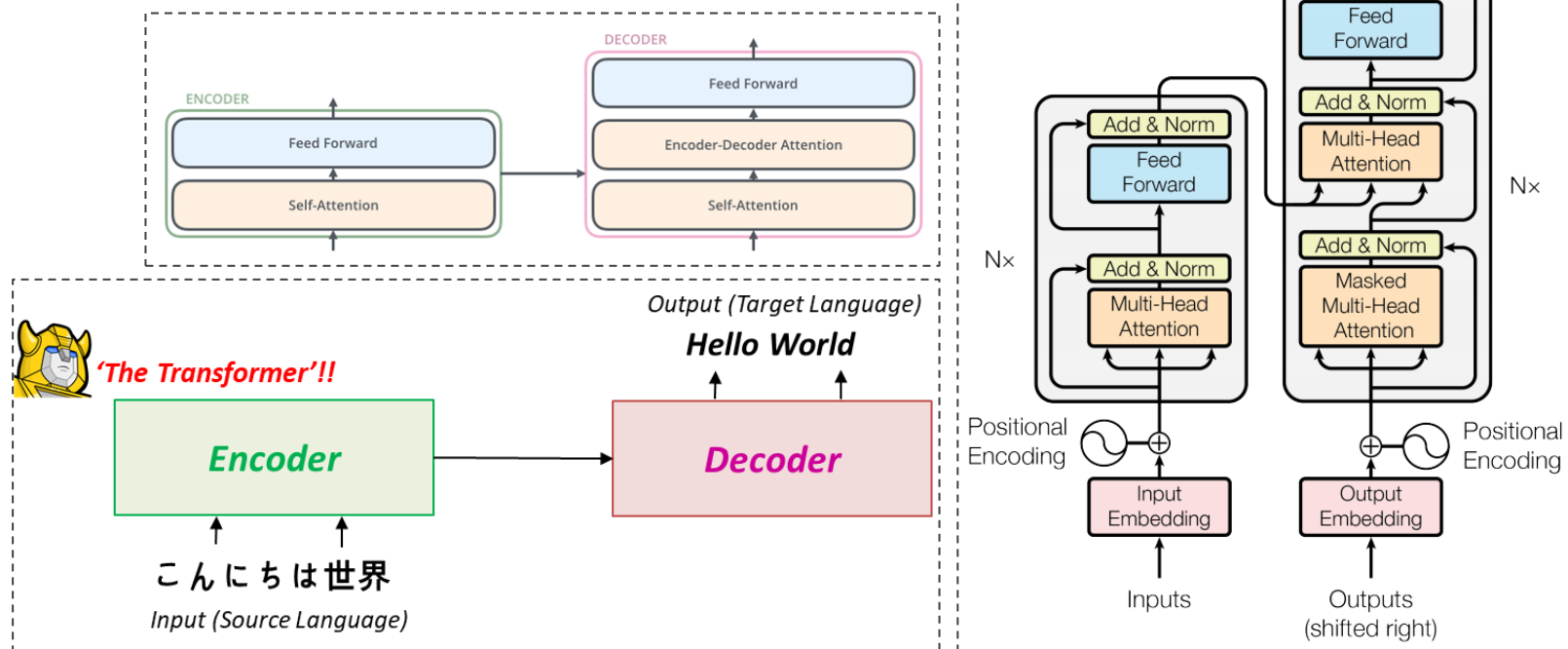


Lecture 11 Summary

Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT

The Rise of the Pre-trained Model

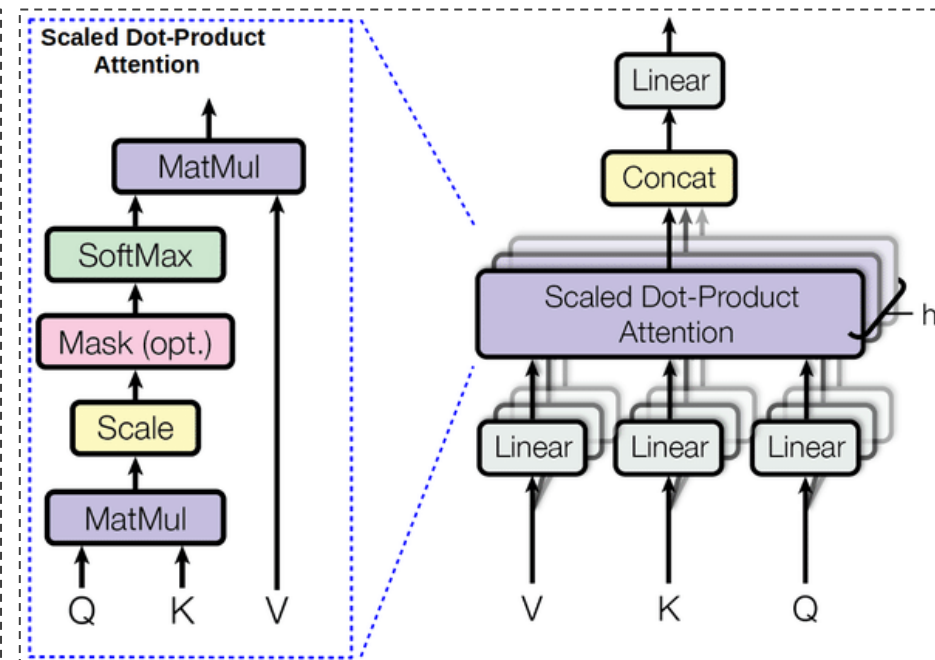
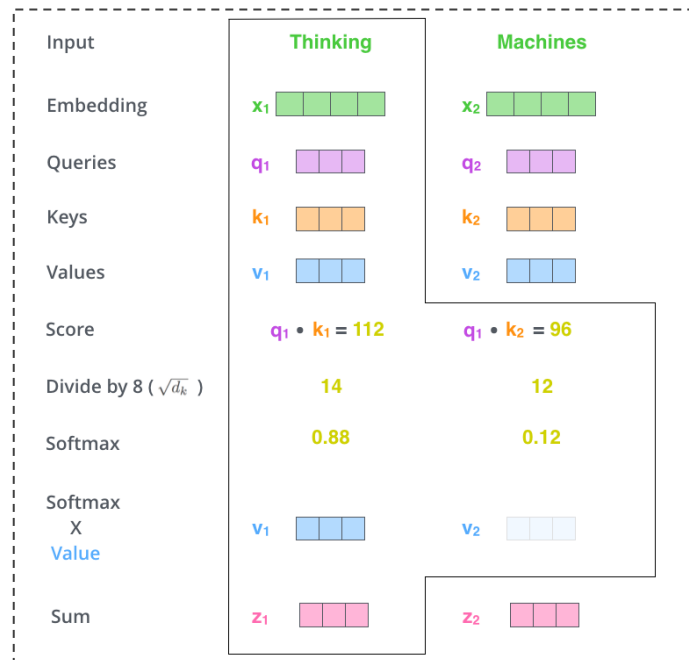
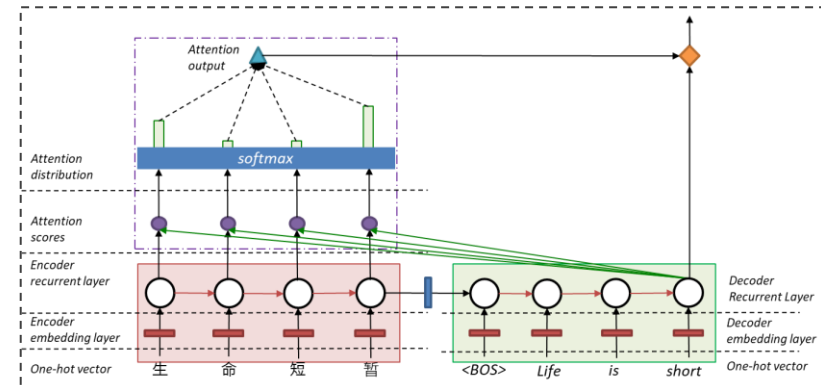


Lecture 11 Summary

Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT

The Rise of the Pre-trained Model

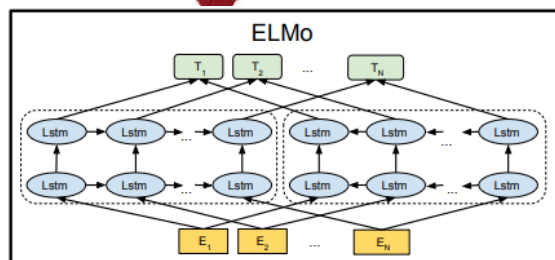


Lecture 11 Summary

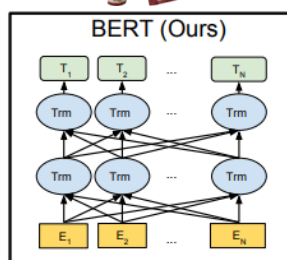
Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT

The Rise of the Pre-trained Model



(Peters et al, 2018)



(Devlin et al, 2018)

Semi-supervised Learning Step

Model:



Dataset:



Objective:

Predict the masked word
(language modeling)

Supervised Learning Step

Classifier

75% Spam
25% Not Spam

Model:
(pre-trained
in step #1)



Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached...	Not Spam

Lecture 12 Summary

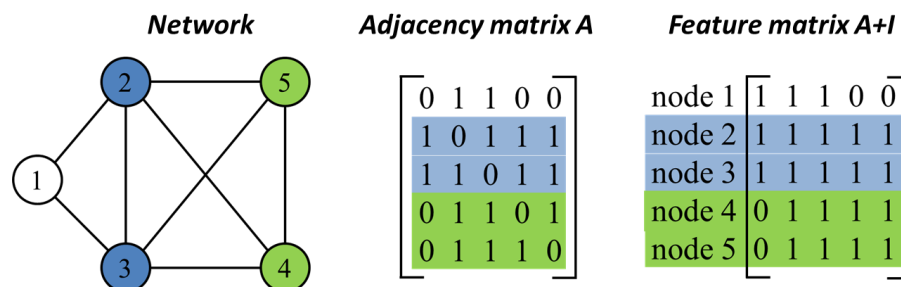
Graph and Natural Language Processing

Graph Neural Networks

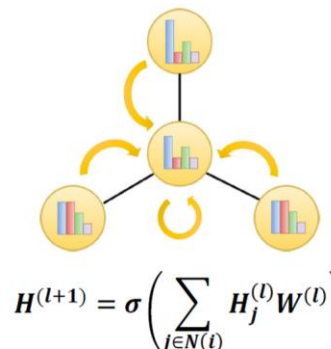
Graph Neural Networks – Task-based

Graph Neural Networks and NLP Application

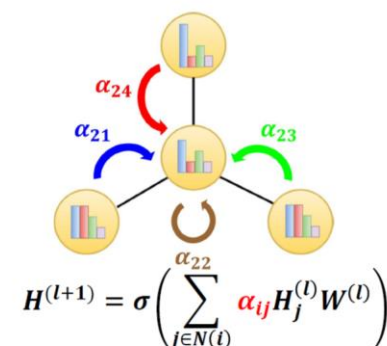
- Text Classification
- Document Timestamping
- Text to image generation



Vanilla GCN updates information of neighbor atoms **with same importance**.



Attention mechanism enables it to update nodes **with different importance**



Lecture 12 Summary

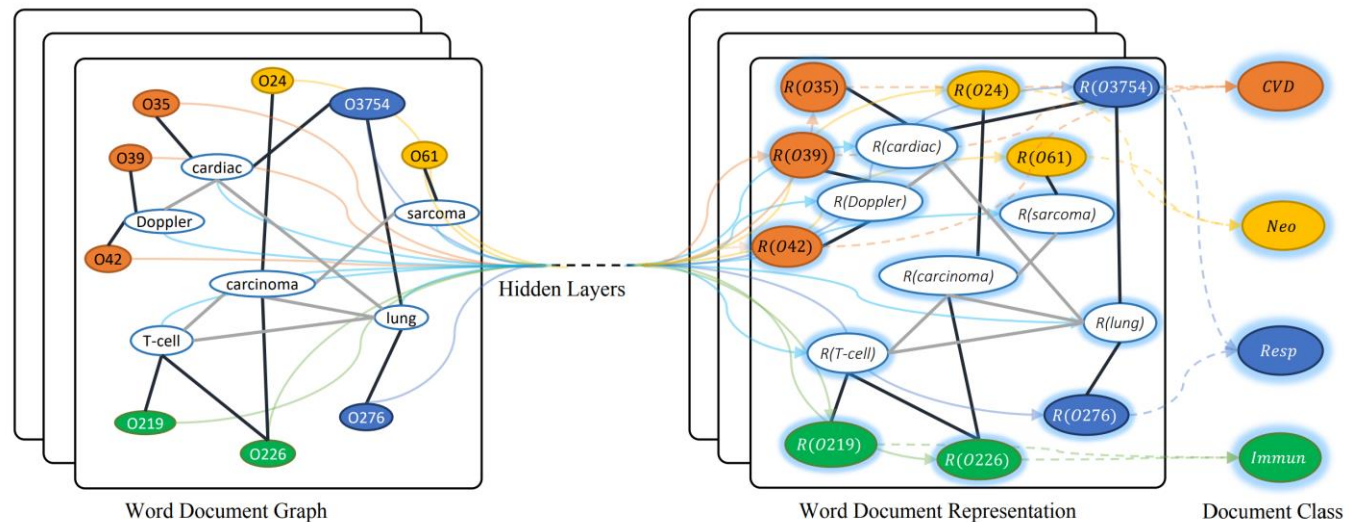
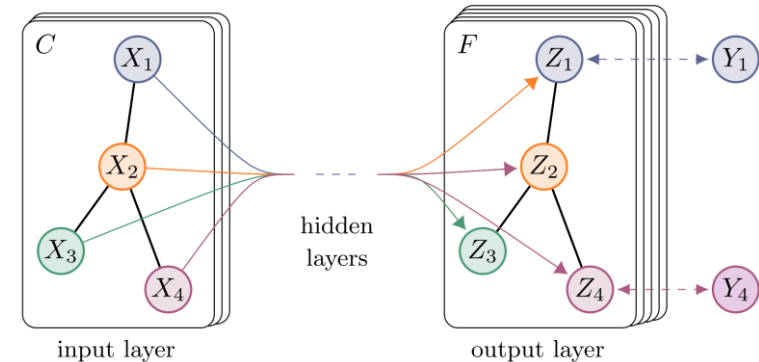
Graph and Natural Language Processing

Graph Neural Networks

Graph Neural Networks – Task-based

Graph Neural Networks and NLP Application

- Text Classification
- Document Timestamping
- Text to image generation



Assessment Update

[Original Assessment Plan]

Assessment	Weight	Due	Length	
Lab exercises	10%	Multiple weeks (By Monday 6pm or Tuesday 6pm)	n/a	To pass UoS, achieve at least 40% (20 out of 50)
Assignment 1	20%	Week 8	n/a	
Assignment 2	20%	Week 14	n/a	
Final exam	50%	Formal Exam Period	2 hours	To pass UoS, achieve at least 40% (20 out of 50)



[Updated Assessment Plan]

Assessment	Weight	Due	Length	
Lab exercises	10%	Multiple weeks (All by Tuesday 6pm)	n/a	To pass UoS, achieve at least 40% (24 out of 60)
Assignment 1	20%	Week 8	n/a	
Assignment 2	30%	Week 14	n/a	
Oral Quiz – Case Study (5Qs interview style)	15%	1 day after the formal exam period via Zoom [the exact date will be announced soon]	10 mins	To pass UoS, achieve at least 40% (16 out of 40)
Take-home Exam (Open Book)	25%	Formal exam period	1 hours	

Sample Q would be demonstrated in Week 13.

Assessment Update

[Original Assessment Plan]

Assessment	Weight	Due	Length	
Lab exercises	10%	Multiple weeks (By Monday 6pm or Tuesday 6pm)	n/a	To pass UoS, achieve at least 40% (20 out of 50)
Assignment 1	20%	Week 8	n/a	
Assignment 2	20%	Week 14	n/a	
Final exam	50%	Formal Exam Period	2 hours	To pass UoS, achieve at least 40% (20 out of 50)



[Updated Assessment Plan]

Assessment	Weight	Due	Length	
Lab exercises	10%	Multiple weeks (All by Tuesday 6pm)	n/a	To pass UoS, achieve at least 40% (24 out of 60)
Assignment 1	20%	Week 8	n/a	
Assignment 2	30%	Week 14	n/a	
Oral Quiz – Case Study (5Qs interview style)	15%	1 day after the formal exam period via Zoom [the exact date will be announced soon]	10 mins	To pass UoS, achieve at least 40% (16 out of 40)
Take-home Exam (Open Book)	25%	Formal exam period	1 hours	

Sample Q would be demonstrated in Week 13.

Final Exam (40%) – Take-home (25%) and Oral (15%)

Take home exam (25%) - Canvas

1:00PM Saturday 20 June 2020 **Check your personal exam timetable at Sydney Student*

Canvas Exam Site

- The canvas exam site is different that the Canvas site we use during the semester for teaching
- It is called Final exam for COMP5046
- The Exams office will give you access to the exam site
- The recommended browsers for Canvas: Chrome and Firefox, not Internet Explorer or others

Exam Conditions

- **1 hour**, Online exam (+10 minutes reading time)
- **Open book** examination, Unsupervised (Un-proctored)
- **Complete ALL questions** in the space provided on the exam paper
- **Eight (8) questions (total 25 marks) = five (5) 2 marks questions + three (3) 5 marks questions**

Final Exam (40%) – Take-home (25%) and Oral (15%)

Oral exam (15%) - Zoom

21-23 June 2020 **The oral exam timetable will be shared on 31st of May*

Exam Conditions

- **10mins**, Zoom-based Interview.
- **Oral Interview Style**

Before the Oral Exam:

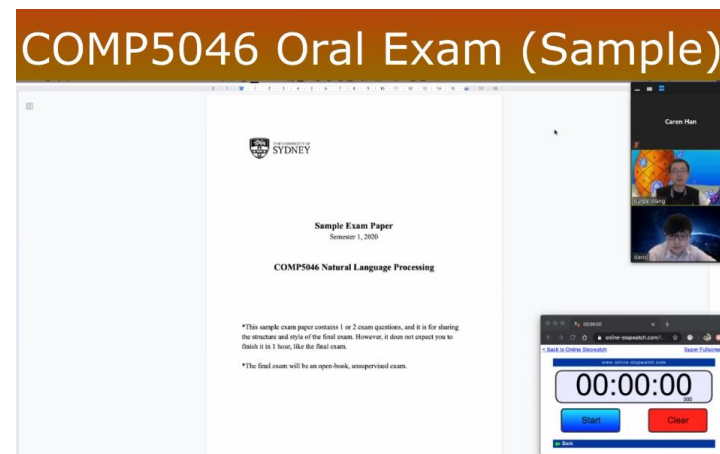
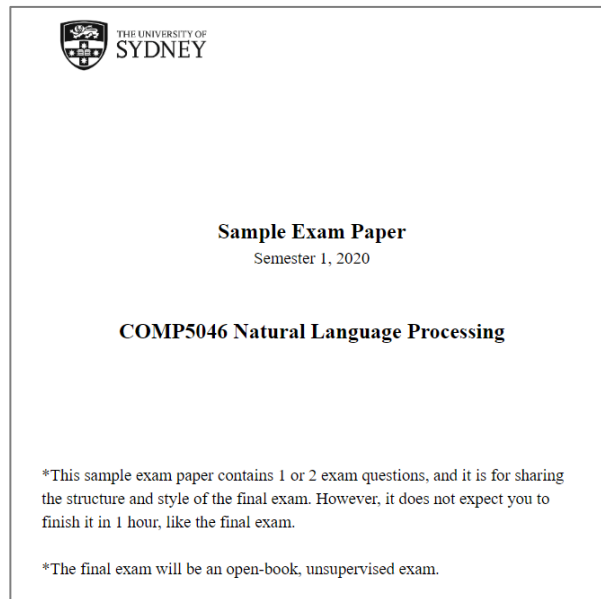
- *Your candidate timetable will tell you when your oral examination is scheduled to begin (The oral exam timetable will be shared on 31st of May)*
- *You will be asked to arrive earlier than the scheduled time for your examination. For all candidates, this will be no later than 10 minutes prior to their examination start time. These arrangements are to ensure the quality of the internet connection.*
- *When you arrive at the room, you will be asked to show your student ID (card)*
- *You will be asked to stand by in the waiting room.*

Final Exam Preview! (All available in Canvas)

Let's Have a Look at this Together!

Sample Online Exam – Question and Answer

Sample Oral Exam – Interview



Reminder! Suggestion!

The questions are not easily googlable.

Your answer will be checked by the plagiarism checking software.

Well organized and be quick

Allocate time according to the marks for each question

Students in different time zones

- Important for international students who could not come to Australia and exchange students who had to return to their country.
- If you are in a different time zone and the exam is too early/too late for you, you can apply for Special Consideration to have special arrangements for the exam:

<https://www.sydney.edu.au/students/special-consideration.html#time-zone>

All Good! Thanks for following and surviving



GOOD LUCK in YOUR FINAL EXAM