EEP 596: AI and Health Care || Lecture 11

Dr. Karthik Mohan

Univ. of Washington, Seattle

May 15, 2022

Last Lecture

1. Handwriting recognition
2. Automated scribing from notes or audio

Digital scribe

topic segmentation

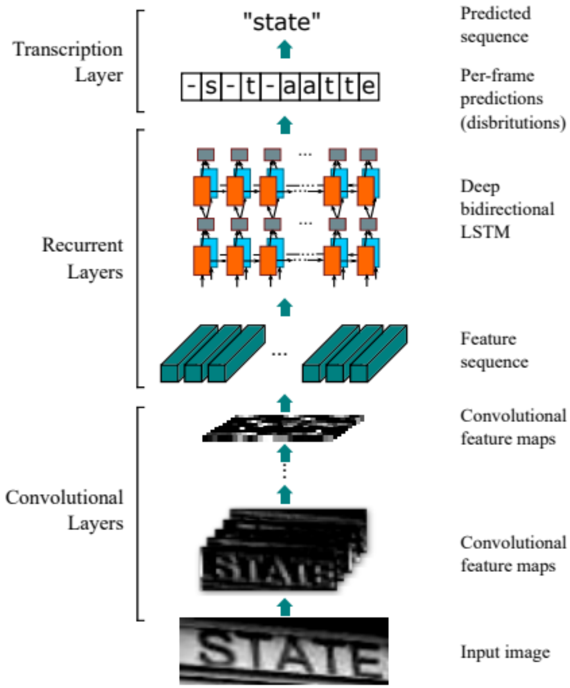
Summarization

Transcribing

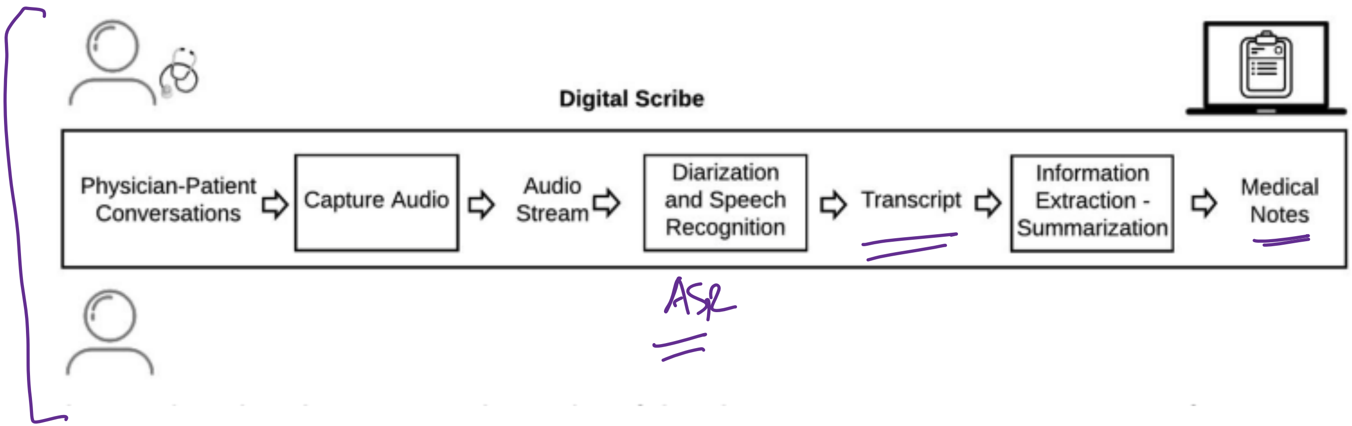
Today

1. Deep Learning for OCR/Handwriting Recognition
2. NLP for summarization, topic segmentation, generating an automated and structured EHR

Deep Learning for OCR



Ideal Digital Scribe



Challenges in Digital Scribing

Table 1. The challenges associated with the various tasks a digital scribe must perform.

Task	Challenge
Recording audio	• High ambient noise
	• Microphone fidelity
	• Multiple speakers
	• Microphone positioning relative to clinician and patient
Automatic speech recognition	• Varying audio quality
	• High ambient noise
	• Multiple speakers
	• Disfluencies, false starts, interruptions, non-lexical pauses
	• Complexity of medical vocabulary
	• Variable speaker volume due to distance to microphone and relative positioning
	• Differentiating multiple speakers in the audio (speaker diarization)
Topic segmentation	• Unstructured conversations
	• Non-linear progression of topics during a medical conversation
Medical concept extraction	• Noisy output of programs mapping text to UMLS
	• Tuning of parameters of tools used to map text to UMLS
	• Contextual inference (understanding the appropriate meaning of a word or phrase given the context)
	• Phenomena in spontaneous speech such as zero anaphora, thinking aloud, topic drift
Summarization	• Summarization of non-verbal unstructured communication
	• Integrating medical knowledge to identify relevant information
	• Contextual inference
	• Resolving conflicting information from the patient
	• Updating hypotheses as the patient discloses more information
	• Generating summaries to train a summarization ML model
Data collection	• Clinician and patient privacy concerns
	• Costly data collection and labeling
	• Patient consent to be audio recorded and use the data for research purposes
	• De-identification and anonymization of data
	• Expensive datasets
	• Data held privately as an intellectual property asset
	• Clinician reluctance to be recorded due to fear of legal liabilities and extra workload

Challenges in Digital Scribing

Conversation	Topic
Dr: Do you drink alcohol? Pt: Occasionally. Dr: Like every week or two Pt: Every two weeks.	Social History
Dr: Ok [pause] Alright and Family history, mum and dad are they alive and well and healthy?	Family History
Dr: So, when you're hungry it feels uncomfortable. Have you had that before or is this all new, do you think?	Present Complaint
Pt: I think last year I used to have that but then I have some medicine to reduce the acid level and then it improved. Yeah. Dr: Is that something that was prescribed in Australia?	Drug History
[The patient steps on the scale.] Dr: So, you haven't lost a whole lot of weight? Pt: It doesn't show already. Do I have to press any buttons? Dr: Ah, yes. Step off. Fifty-four point three. [The patient steps off the scale.] Pt: So, already one kilo decrease. Dr: Okay. One kilo less? Pt: Yeah.	Physical Examination
Dr: Any blood in the bowel motions or vomiting blood, nothing like that? Pt: No, no.	Present Complaint

Fig. 2 Three examples of transitions of clinician–patient conversations lacking clear boundaries and structure. Medical conversation fragments are on the left and the respective topics are on the right. Medical conversations do not appear to follow a classic linear model of defined information seeking activities. The nonlinearity of activities requires digital scribes to link disparate information fragments, merge their content, and abstract coherent information summaries.

Today's Lecture

1. DL for OCR
2. NLP methods for ~~extracting topic segmentation~~ Summarization and generating an automated and structured EHR

Sequence structure in NLP

Example

I love this car! Positive Sentiment

Sequence structure in NLP

Example

I love this car!   Positive Sentiment

-ve   +ve

Example

I am not sure I love this car!   Negative Sentiment

Sequence structure in NLP

Example

I love this car! Positive Sentiment

Example

I am not sure I love this car! Negative Sentiment

Example

I don't think its a bad car at all! $\rightarrow$ Positive Sentiment

Sequence structure in NLP

Example

I love this car! Positive Sentiment

Example

I am not sure I love this car! Negative Sentiment

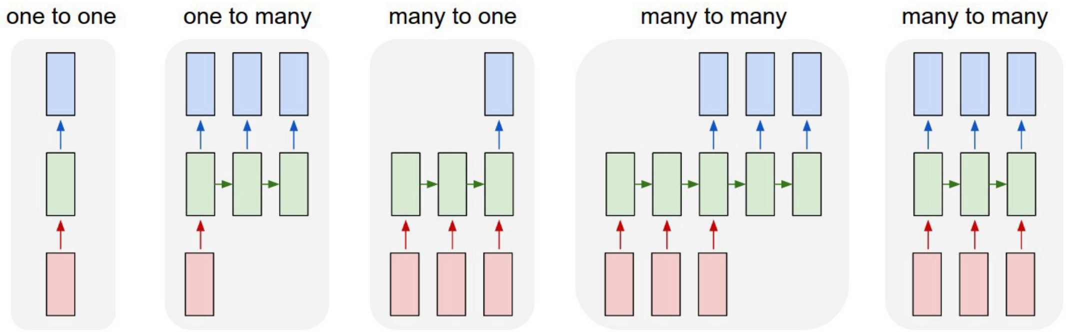
Example

I don't think its a bad car at all! $\rightarrow$ Positive Sentiment

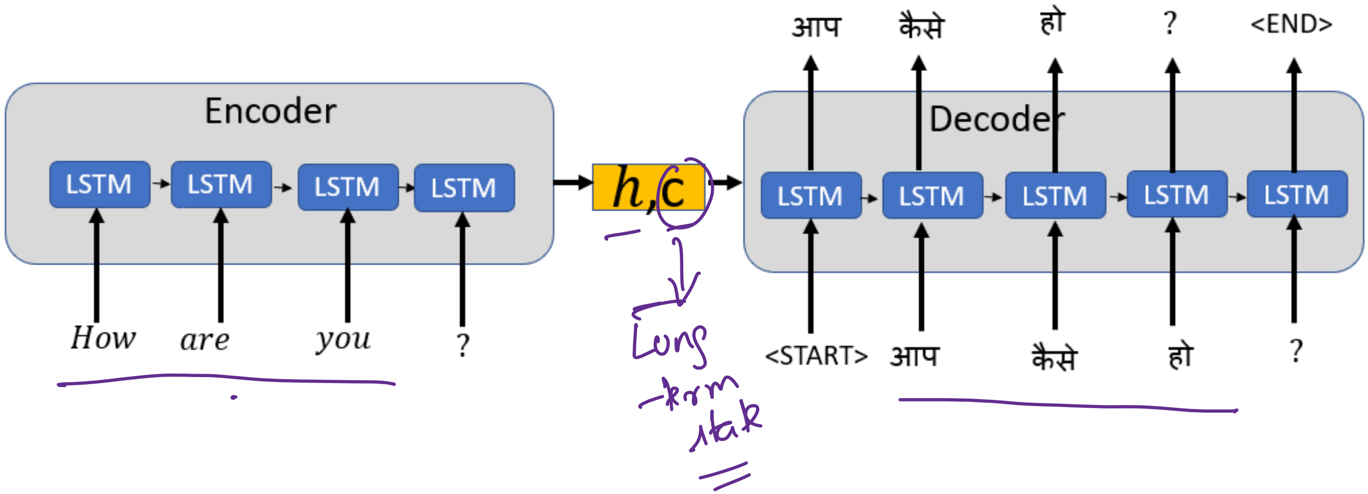
Example

Have to carry the **context(state)** from some-time back to fully understand what's happening!

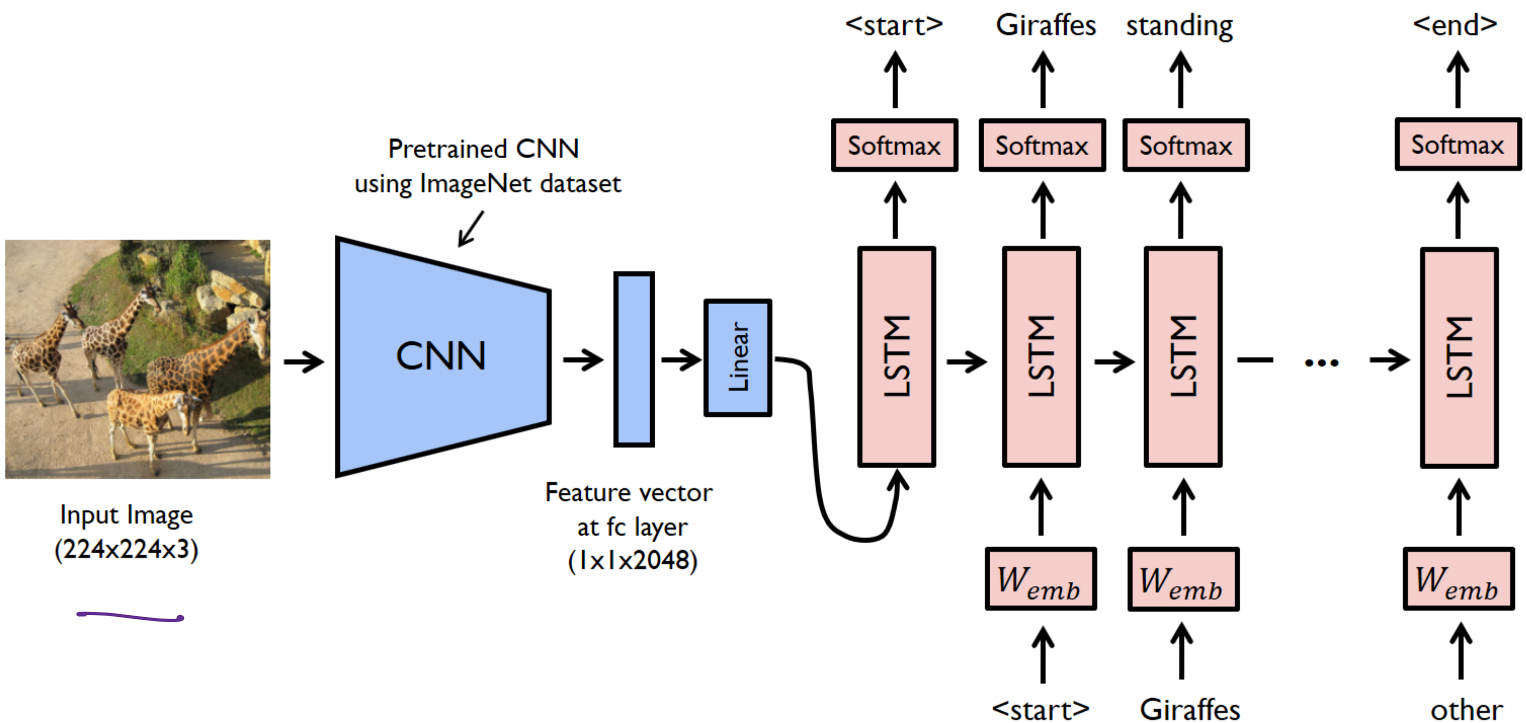
Sequence to Sequence Model (LSTM) Applications



Sequence to Sequence Model (LSTM) Applications



Sequence to Sequence Model (LSTM) Applications



Applications of NLP for Digital Scribing

Applications

1. Topic Modeling/Topic Segmentation

Applications of NLP for Digital Scribing

Applications

1. Topic Modeling/Topic Segmentation
2. Notes/Document Summarization

Applications of NLP for Digital Scribing

Applications

1. Topic Modeling/Topic Segmentation
2. Notes/Document Summarization
3. Chat bots
4. Entity Extraction

Applications of NLP for Digital Scribing

Applications

1. Topic Modeling/Topic Segmentation
2. Notes/Document Summarization
3. Chat bots
4. More?

Genetics, Life, Computation

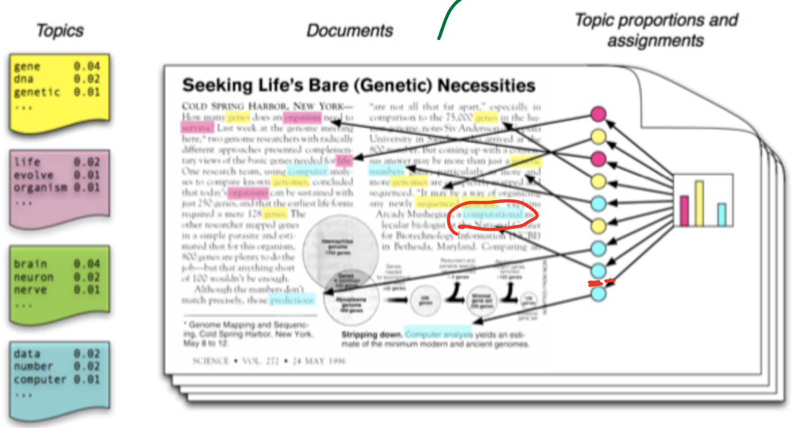


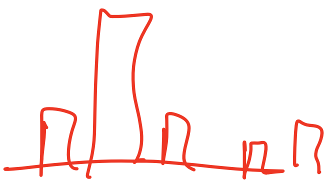
Figure source: Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.

LDA

Latent

Dirichlet

Allocation



WH Speaker . . . President Topic → Prob distn. over words (White House, Speaker, President, . . .)

Topic (Politics)

Document → Prob distn. over topics

Topic Modeling vs Topic Segmentation

ICE #0

Which of the following statements are true:

a. They refer to the same set of techniques
b. Topic segmentation deals with segregating sentences into topics while topic modeling gives overall understanding of topics in a document
c. Topic Modeling can be used to do topic segmentation
d. Topic segmentation tells us how many topic segments exist in a document

Document Summarization

Fully Abstractive Summarization → Abstractive Extractive Summary

So far none of the videos were ...

Input Article

Marseille, France (CNN) The French prosecutor leading an investigation into the crash of Germanwings Flight 9525 insisted Wednesday that he was not aware of any video footage from on board the plane. Marseille prosecutor Brice Robin told CNN that " so far no videos were used in the crash investigation . " He added, " A person who has such a video needs to immediately give it to the investigators . " Robin\'s comments follow claims by two magazines, German daily Bild and French Paris Match, of a cell phone video showing the harrowing final seconds from on board Germanwings Flight 9525 as it crashed into the French Alps . All 150 on board were killed. Paris Match and Bild reported that the video was recovered from a phone at the wreckage site. ...

Text Summarization Models

- 2 Abstractive summarization
- 1 Extractive summarization

Generated summary

Prosecutor : " So far no videos were used in the crash investigation "

Extractive summary

marseille prosecutor brice robin told cnn that " so far no videos were used in the crash investigation . " robin \'s comments follow claims by two magazines , german daily bild and french paris match , of a cell phone video showing the harrowing final seconds from on board germanwings flight 9525 as it crashed into the french alps . paris match and bild reported that the video was recovered from a phone at the wreckage site .

Document Summarization — Extractive

Baseline model:-

1) News Documents → Top 3 sentences

2) Clustering + Representatives
↳ Topic modeling

3) Supervised modeling

Model

Document

Top-K scored sentences
↓
Summary

Submodular / Topic
Diversification

Qualities of a good summary

1) Relevant & Important ✓
2) Coverage — Diversity
3) Repetition of information

Evaluation Metrics

1. ROUGE score: Recall-Oriented Understudy for Gisting Evaluation
2. ROUGE-N: N-gram overlap between two summaries (expressed as a fraction or percentage)

ICE #1

ROUGE-1

Consider the truth summary and an automated summary of an article on document summarization of medical documents ! Find the ROUGE-N score based on finding the proportion of N-grams in the truth summary that are also in the automated summary for $N = 1$.

Truth Summary: The paper discusses thoroughly the promising paths for future research in medical documents summarization. It mainly focuses on the issue of scaling to large collections of documents in various languages and from different media.

Automated Summary: This paper discusses summarization for medical documents including issues of scaling to large collections.

ROUGE-1 $=$

a) 0.31 b) 0.25 c) 0.38 d) 0.45

Reference paper!

Rouge Metrics

Variations

Rouge-N where $N = 1, 2, L$.

Recall (Default), Precision, F-score

Extractive

Abstractive

→ 0.9

→ 0.2

Rouge 1

Rouge 5

ICE #2

ROUGE-1 precision

Consider the truth summary and an automated summary of an article on document summarization of medical documents ! Find the ROUGE-N precision score based on finding the proportion of N-grams in the truth ~~automated~~ summary that are also in the automated ~~truth~~ summary for $N = 1$.

Truth Summary: The paper discusses thoroughly the promising paths for future research in medical documents summarization. It mainly focuses on the issue of scaling to large collections of documents in various languages and from different media.

Automated Summary: This paper discusses summarization for medical documents including issues of scaling to large collections. "stopword"

ROUGE-1 precision = $8/9$?

a) 0.8 b) 0.86 c) 0.92 d) 0.98

Reference paper!

Extractive	Abstractive
- Picking exact sentences from a doc	- Free flow summarization
- May suffer from flow	- Suffer from hallucinations "This document document describ... - - "
- Relevant	- Issues with relevance
- Less natural	- More natural

Extractive - Abstractive Summarization

↳ Best of both worlds

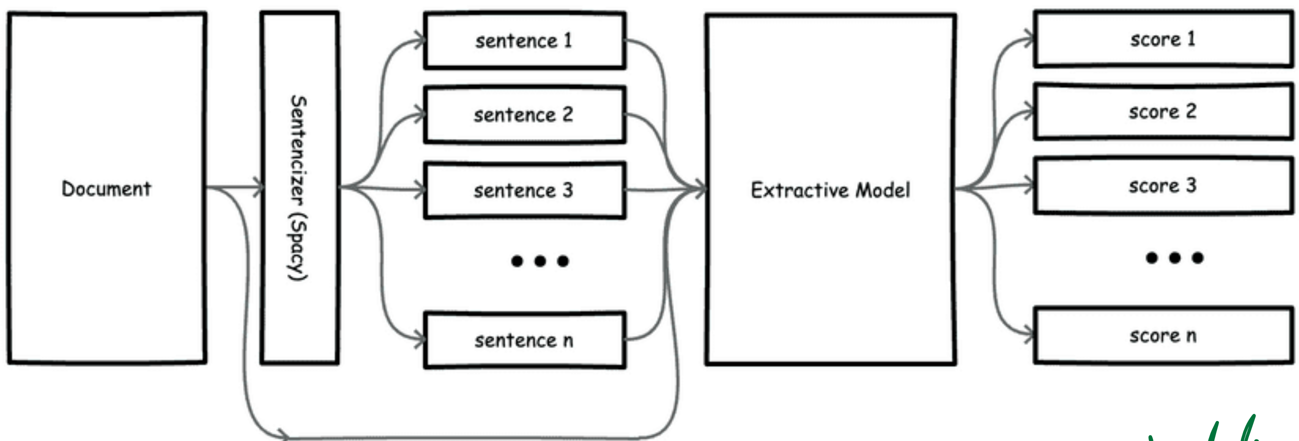
Document Summarization

Document to be summarized is converted to sentences using a Spacy Senticizer.

BERT based model generates a score for each sentence in the document.

Input: pairs of sentence and document
Output: sentence scores

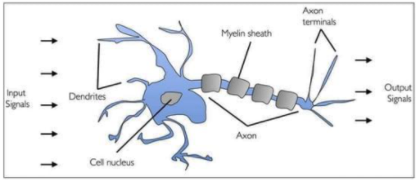
After scoring, we can reorder sentences based on scores, order of appearance (or other post processing criteria), and take top_k sentences as summary

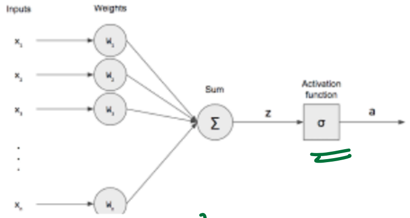


[SBERT → Sentence Embeddings
RoBERTa, BART, .. (BERT)

Evolution of DNN architectures for NLP!

Perceptron



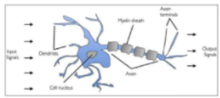


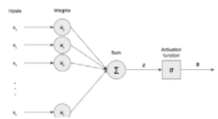
Linear

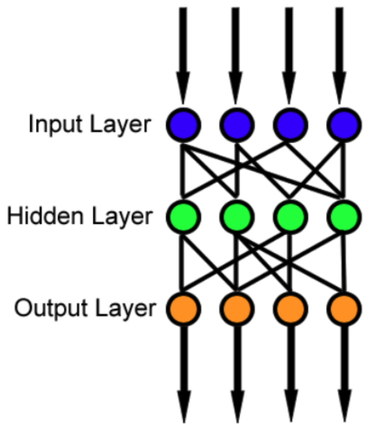
Evolution of DNN architectures for NLP!

Perceptron

Feed Forward NN



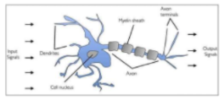


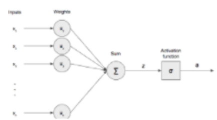


← Non-Linear

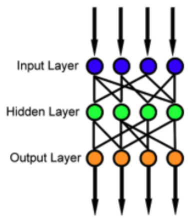
Evolution of DNN architectures for NLP!

Perceptron

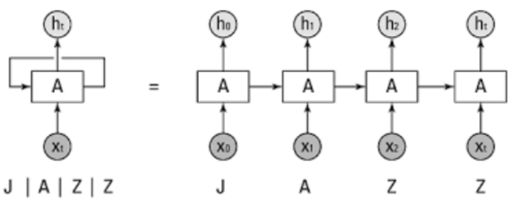




Feed Forward NN



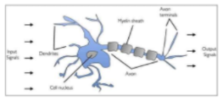
RNN



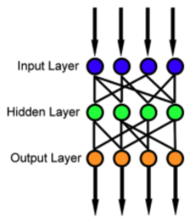
word-word

Evolution of DNN architectures for NLP!

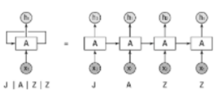
Perceptron



Feed Forward NN



RNN

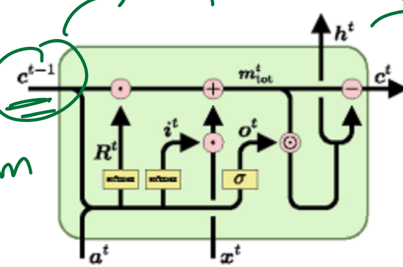


LSTM

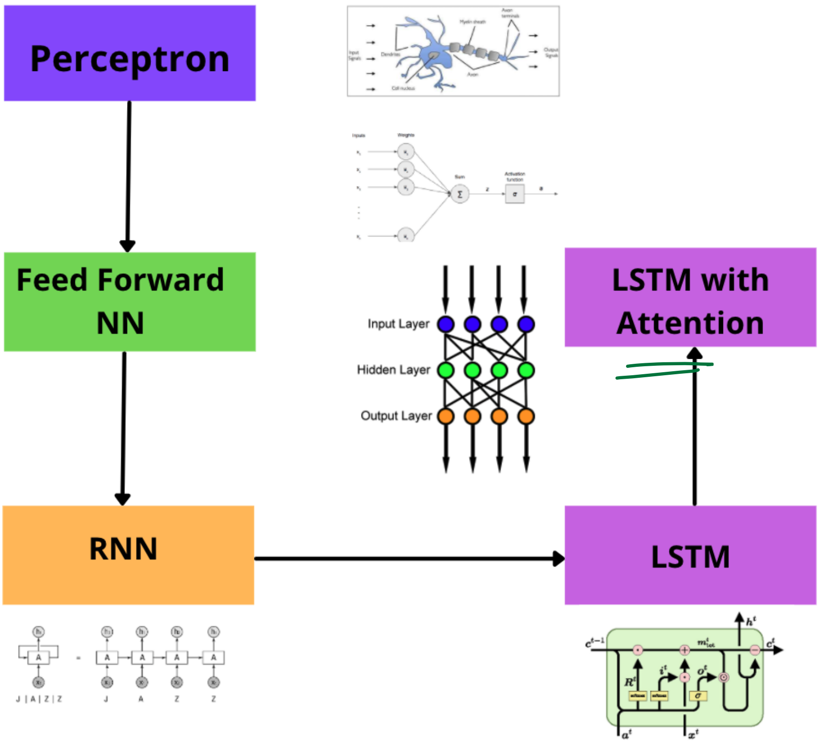
Long Short Term Memory

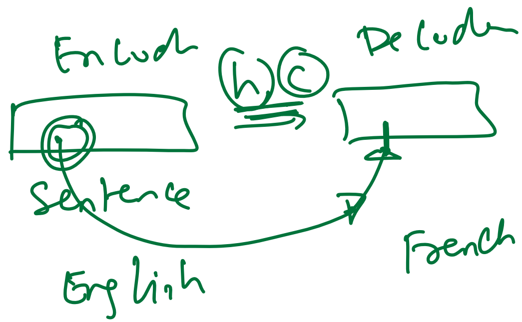
Long term state

Immediate state (RNN)

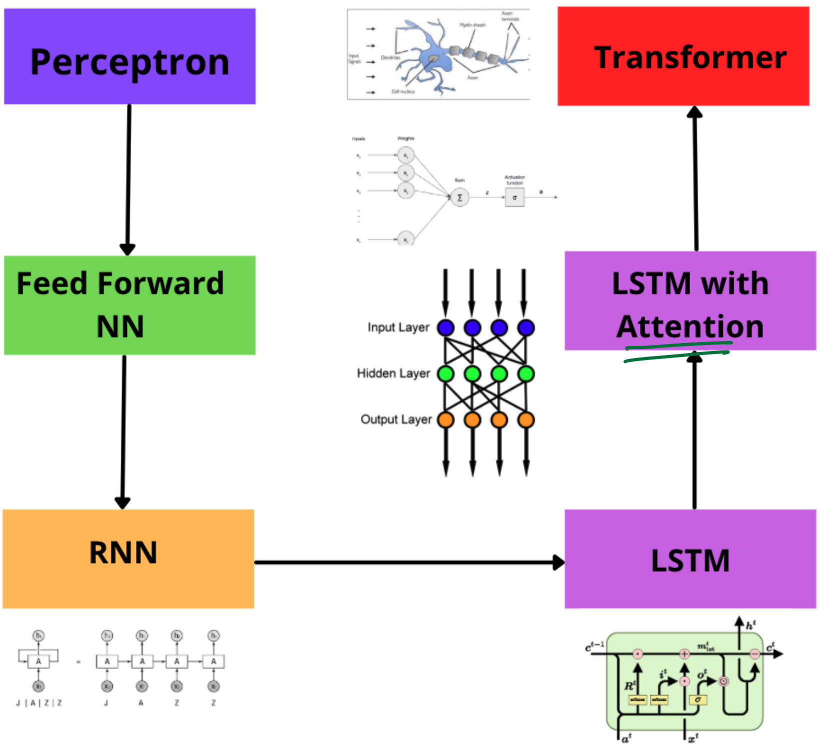


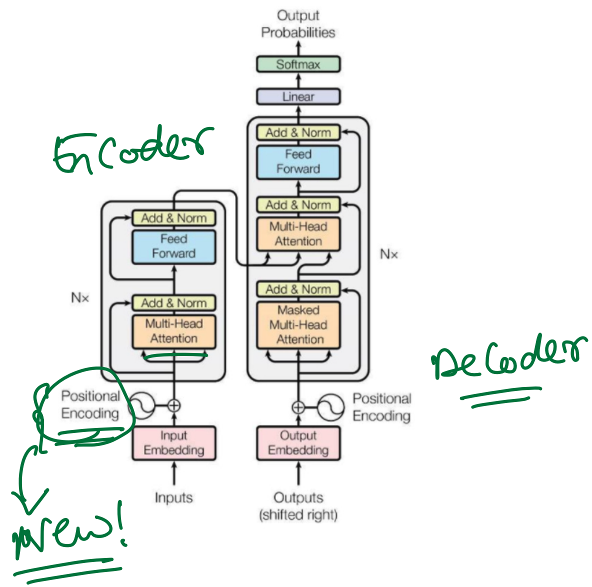
Evolution of DNN architectures for NLP!



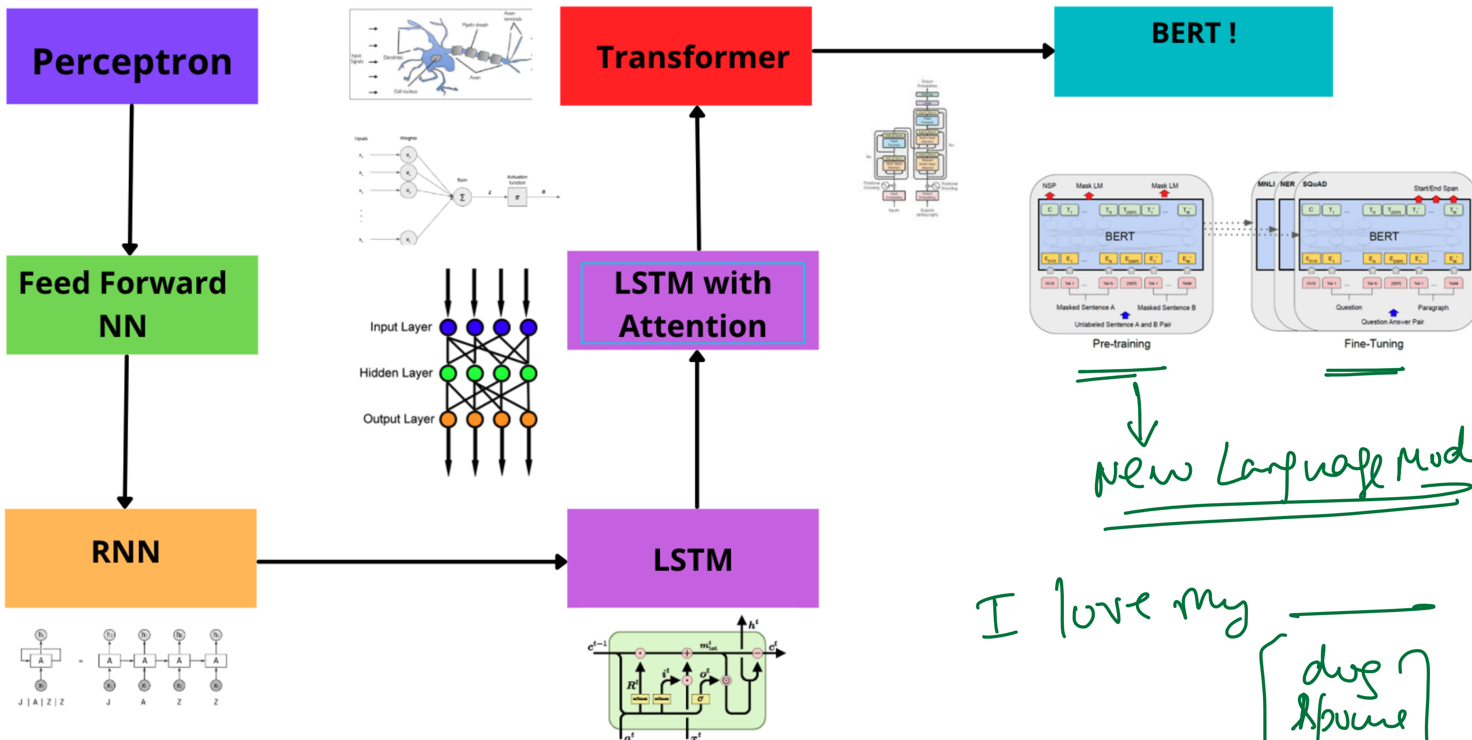


Evolution of DNN architectures for NLP!





Evolution of DNN architectures for NLP!

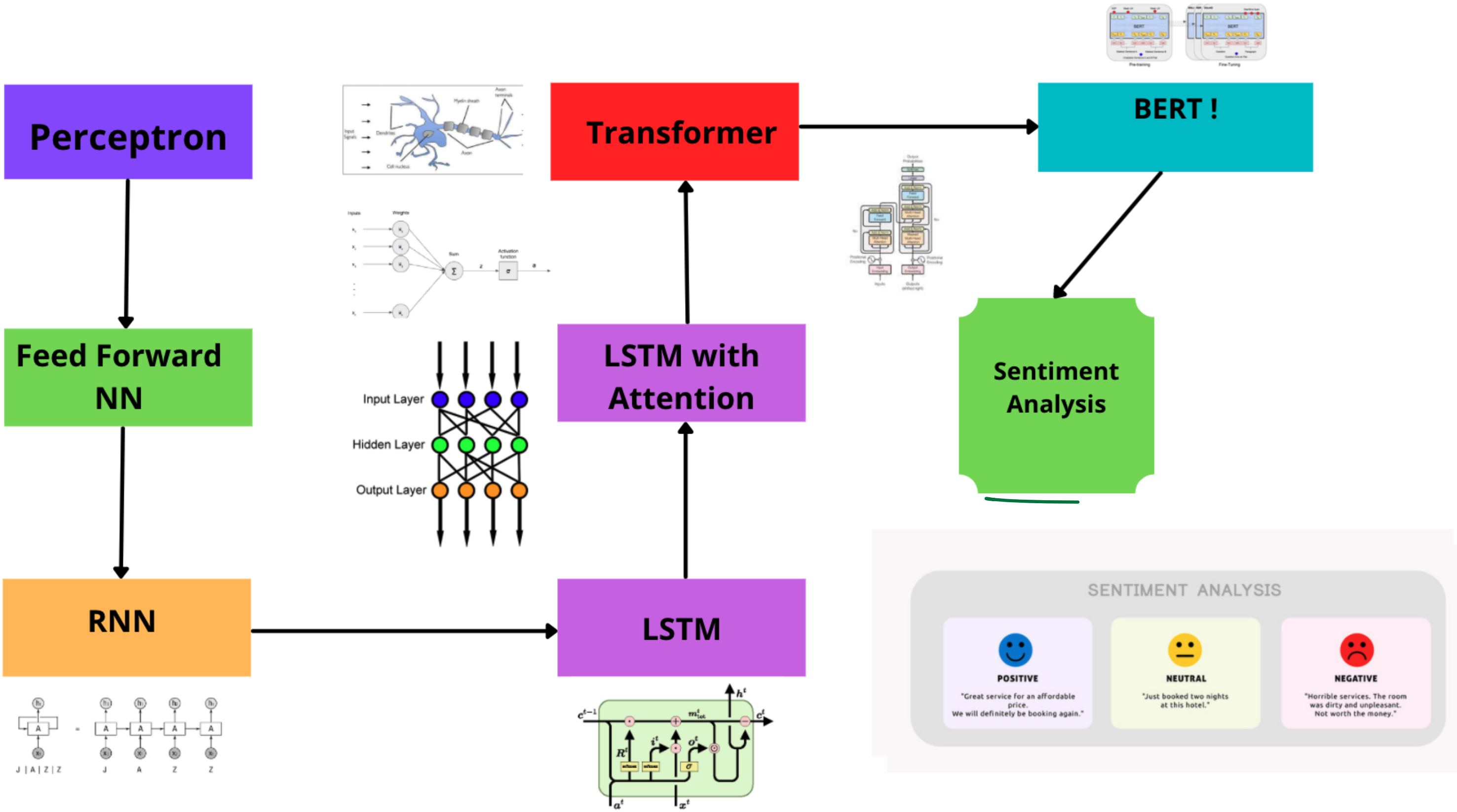


new Language model!

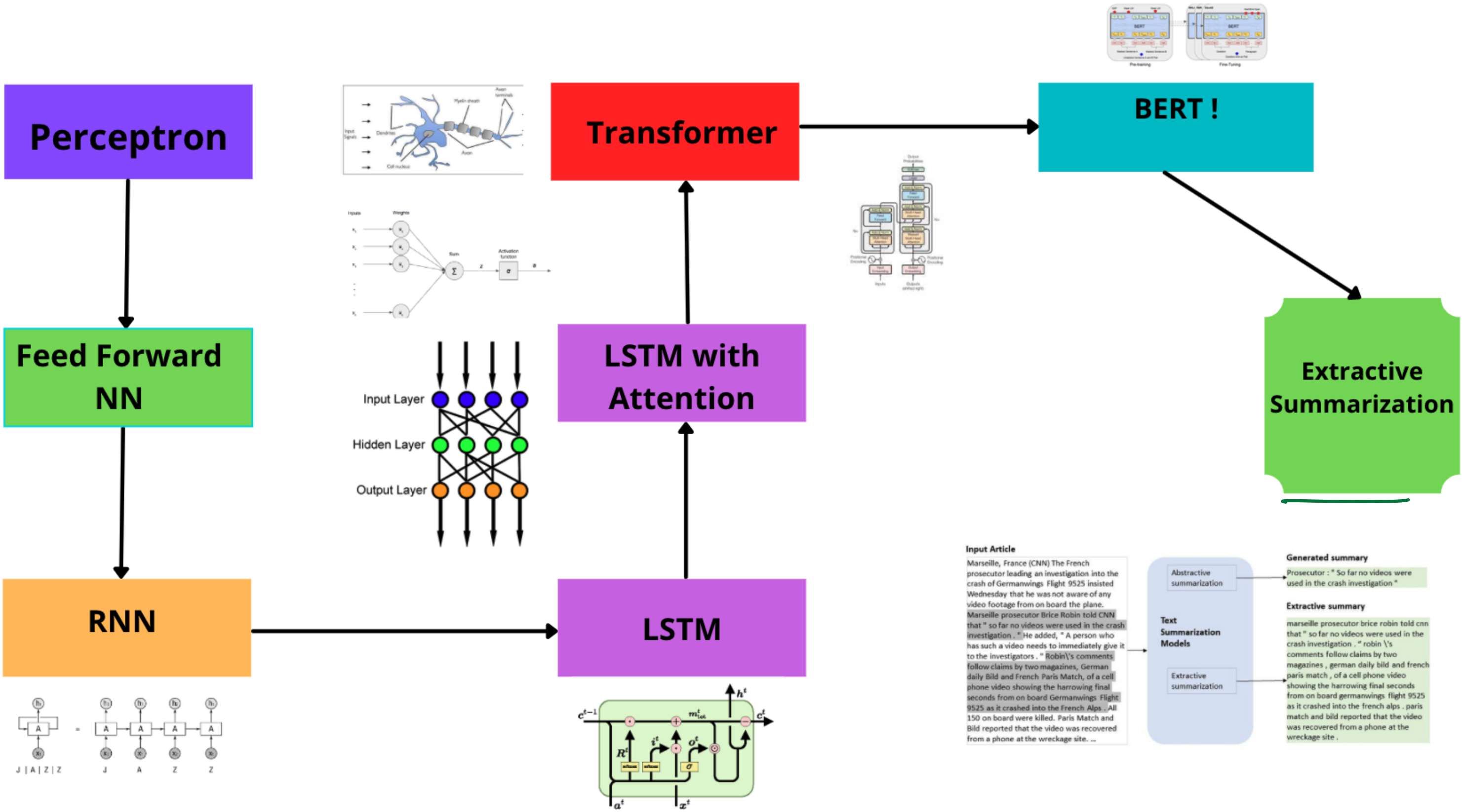
I love my ____

dog
Spouse
:

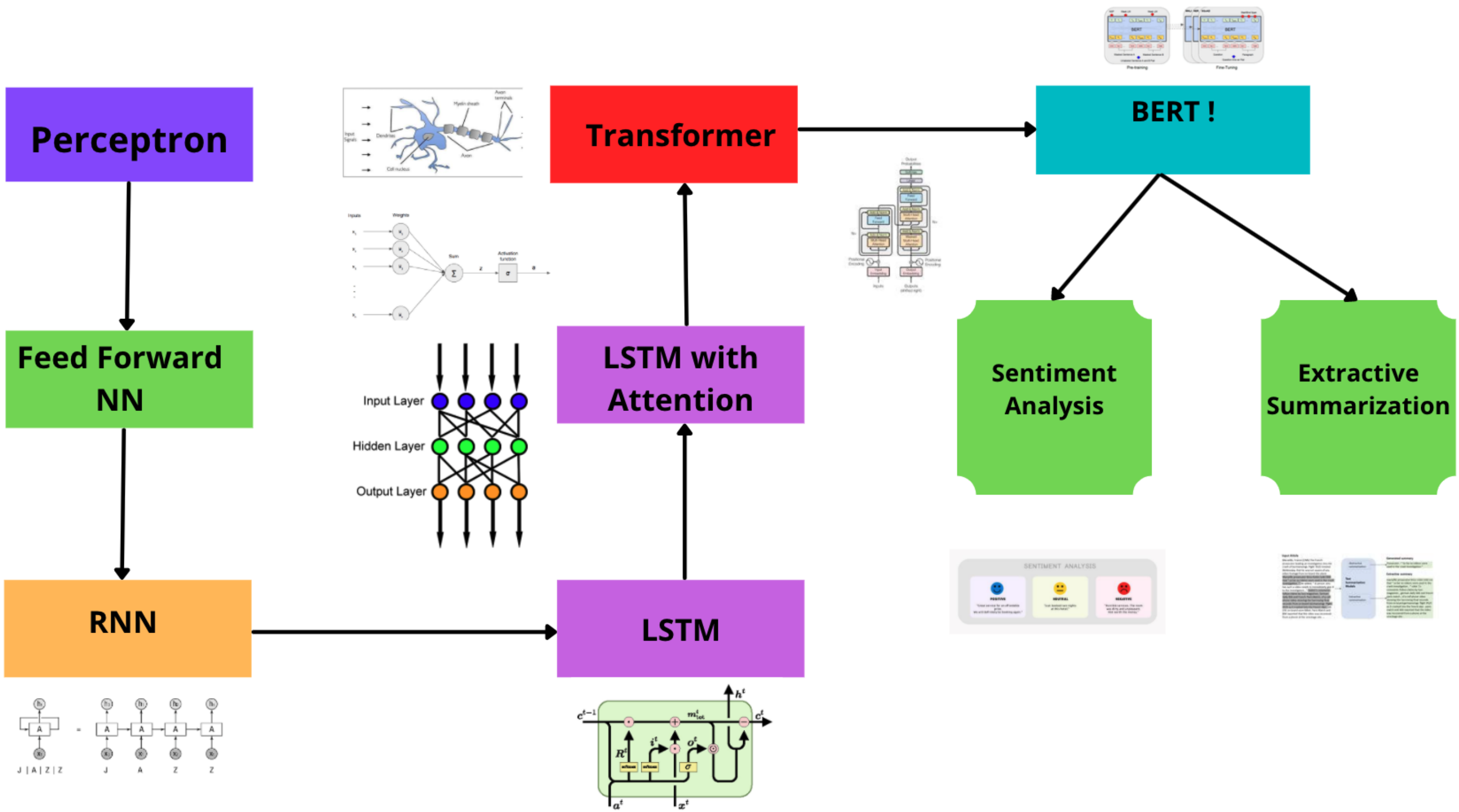
Evolution of DNN architectures for NLP!



Evolution of DNN architectures for NLP!



Evolution of DNN architectures for NLP!



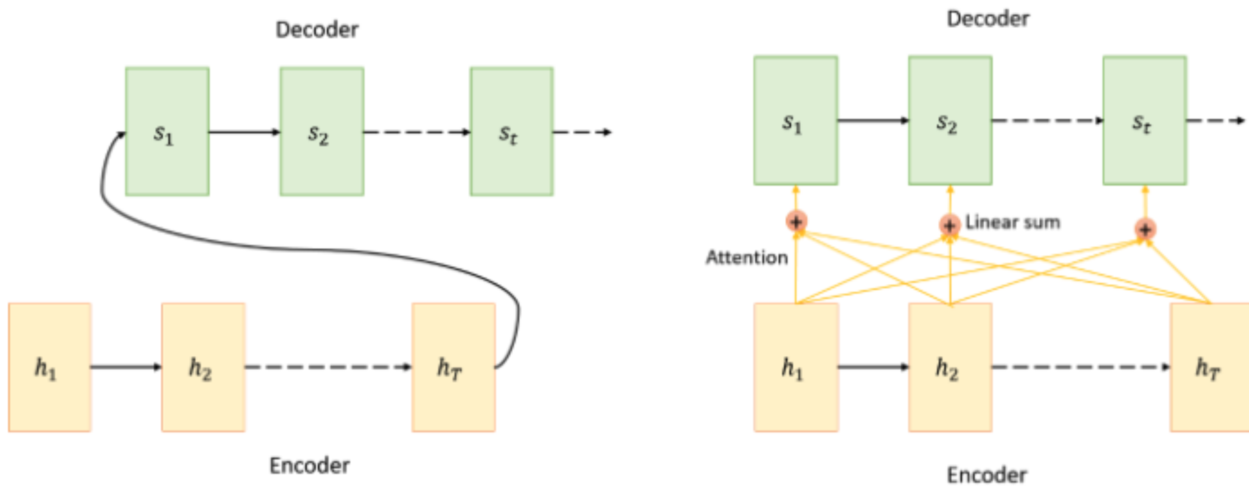
ICE #3

RNN vs LSTM

Which of the following statements are NOT true?

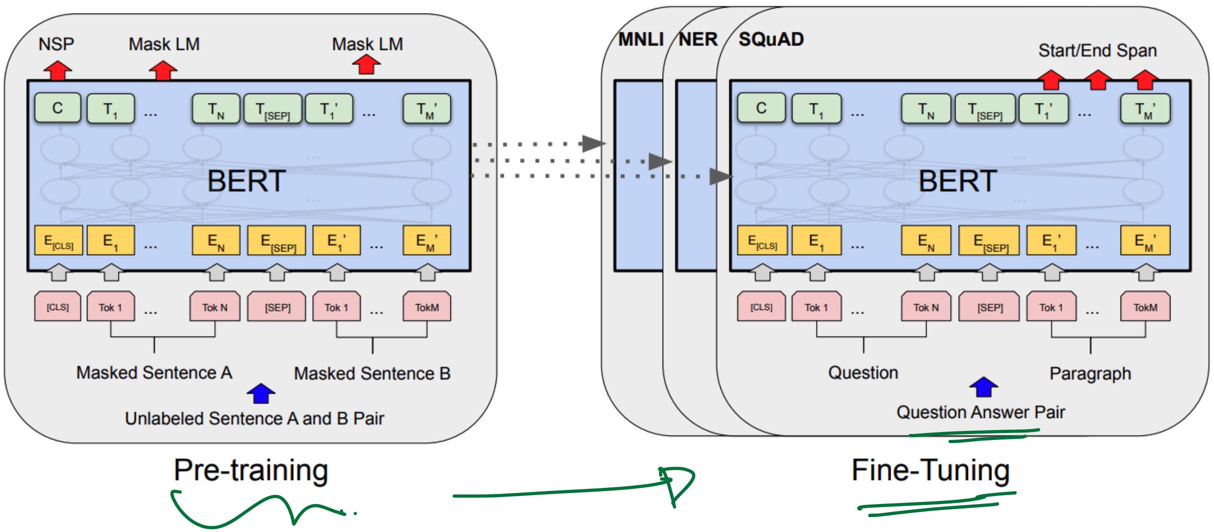
1. LSTM doesn't have the exploding/vanishing gradients issue as it occurs in RNNs
2. LSTM applies to sequential language tasks while RNNs applies to non-sequential language tasks
3. LSTM is better than RNN in most language tasks
4. LSTMs can be used for machine translation tasks

LSTM with attention

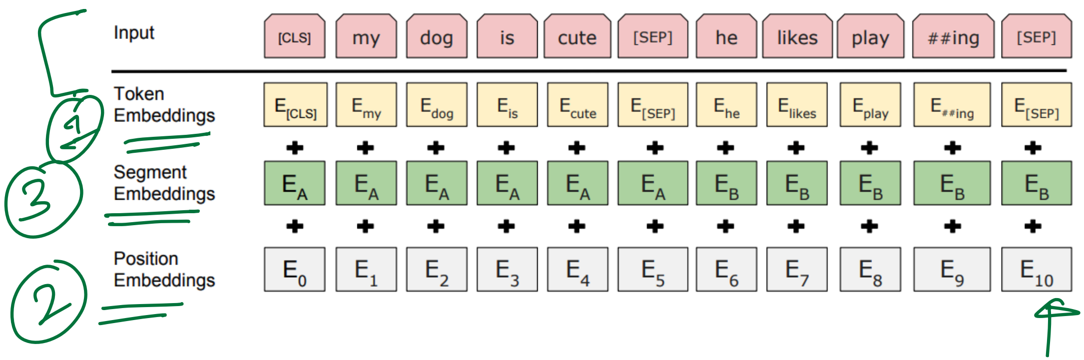


(a) Vanilla Encoder Decoder Architecture **(b) Attention Mechanism**

BERT - Bi-directional Encoders from Transformers



BERT Embeddings



$\rightarrow \mathbb{R}^{256}$ $\mathbb{R}^{64}$

BERT pre-training

Two Tasks

New

1. **Masked LM Model:** Mask a word in the middle of a sentence and have BERT predict the masked word
2. **Next-sentence prediction:** Predict the next sentence - Use both positive and negative labels. How are these generated?

I love my ____ and also my ____

spouse

dog

BERT pre-training

Two Tasks

1. **Masked LM Model:** Mask a word in the middle of a sentence and have BERT predict the masked word
2. **Next-sentence prediction:** Predict the next sentence - Use both positive and negative labels. How are these generated?

ICE #4: Supervised or Un-supervised?

1. Are the above two tasks supervised or un-supervised?

BERT pre-training

Two Tasks

1. **Masked LM Model:** Mask a word in the middle of a sentence and have BERT predict the masked word
2. **Next-sentence prediction:** Predict the next sentence - Use both positive and negative labels. How are these generated?

ICE #4: Supervised or Un-supervised?

1. Are the above two tasks supervised or un-supervised?

Data set!

English Wikipedia and book corpus documents!

BERT - Bi-directional Encoders from Transformers

Data sets

Models

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	**Average**
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**92.7**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**82.1**

LSTM model

BERT - Bi-directional Encoders from Transformers

System	Dev	Test
ESIM+GloVe	51.9	52.7
ESIM+ELMo	59.1	59.2
OpenAI GPT	-	78.0
$\text{BERT}_{\text{BASE}}$	81.6	-
$\text{BERT}_{\text{LARGE}}$	**86.6**	**86.3**
Human (expert)†	-	85.0
Human (5 annotations)†	-	88.0

Table 4: SWAG Dev and Test accuracies. †Human performance is measured with 100 samples, as reported in the SWAG paper.

4 Candidate sentences

Input sentence

ICE #5

MLM

What's the real point of using masked language models (MLM) as compared to regular language models (LM). Select ones that apply!

1. MLMs are used to learn how words fit together in a sentence
2. MLMs incorporate context from both directions and hence lead to better embeddings and predictions as compared to LMs
3. MLMs are great for complicated language tasks such as QA where you need to understand the sentence as a whole to give an appropriate answer to a question

Challenges in Digital Scribing

Conversation	Topic
Dr: Do you drink alcohol? Pt: Occasionally. Dr: Like every week or two Pt: Every two weeks.	Social History
Dr: Ok [pause] Alright and Family history, mum and dad are they alive and well and healthy?	Family History
Dr: So, when you're hungry it feels uncomfortable. Have you had that before or is this all new, do you think? Pt: I think last year I used to have that but then I have some medicine to reduce the acid level and then it improved. Yeah.	Present Complaint
Dr: Is that something that was prescribed in Australia?	Drug History
[The patient steps on the scale.] Dr: So, you haven't lost a whole lot of weight? Pt: It doesn't show already. Do I have to press any buttons? Dr: Ah, yes. Step off. Fifty-four point three. [The patient steps off the scale.] Pt: So, already one kilo decrease. Dr: Okay. One kilo less? Pt: Yeah.	Physical Examination
Dr: Any blood in the bowel motions or vomiting blood, nothing like that? Pt: No, no.	Present Complaint

Fig. 2 Three examples of transitions of clinician–patient conversations lacking clear boundaries and structure. Medical conversation fragments are on the left and the respective topics are on the right. Medical conversations do not appear to follow a classic linear model of defined information seeking activities. The nonlinearity of activities requires digital scribes to link disparate information fragments, merge their content, and abstract coherent information summaries.

Medical Notes Summarization $\rightarrow$ Automated EHR

Requirements and Recipe

1. Summary needs to be structured. Structuring summaries through topics can help. E.g. Past medical history, Current complaint, Past medication, Diagnosis, Next Steps

Medical Notes Summarization → Automated EHR

Requirements and Recipe

1. Summary needs to be structured. Structuring summaries through topics can help. E.g. Past medical history, Current complaint, Past medication, Diagnosis, Next Steps
2. Topic segmentation can be used to identify sentences that are candidates for each topic summary.

Medical Notes Summarization → Automated EHR

Requirements and Recipe

1. Summary needs to be structured. Structuring summaries through topics can help. E.g. Past medical history, Current complaint, Past medication, Diagnosis, Next Steps
2. Topic segmentation can be used to identify sentences that are candidates for each topic summary.
3. For each topic, candidate sentences can go through a summarization model to obtain a summary] BERT

Medical Notes Summarization → Automated EHR

Requirements and Recipe

1. Summary needs to be structured. Structuring summaries through topics can help. E.g. Past medical history, Current complaint, Past medication, Diagnosis, Next Steps
2. Topic segmentation can be used to identify sentences that are candidates for each topic summary.
3. For each topic, candidate sentences can go through a summarization model to obtain a summary
4. Special consideration to preserve 'critical medical observations' in the topic summaries.

BART

References

1. Summarization from medical documents: a survey
2. Abstractive Summarization of Long Medical Documents with Transformers