

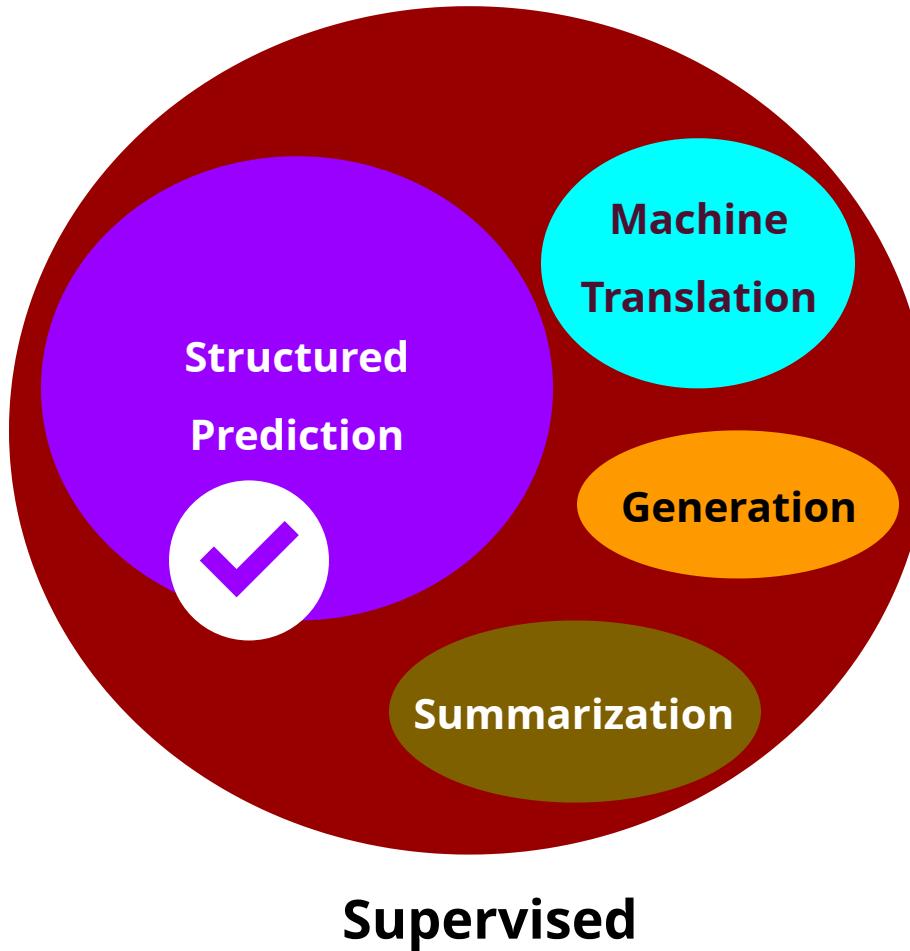
50.040

Natural Language Processing

Lu, Wei



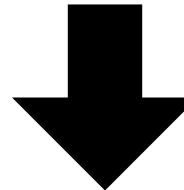
Tasks in NLP



Language Generation



Financial report generation

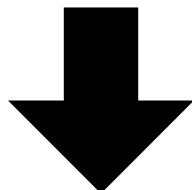


Google has averaged around 25% real growth rate annually, with the exception of 2009 when it was around 7%....

Language Generation



Weather report
generation



Maximum temperature today near 86 degrees. A partly cloudy and warm day is expected. Lowest relative humidity near 33 percent. Expect 13 hours of sunshine ...

Language Generation

Classic Pipeline



Language Generation



Select the information
to be communicated to the user

Language Generation



Decide how the information will be
split among sentences and paragraphs

Language Generation



Generate sentences that are
grammatically correct

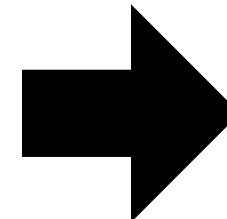
Language Generation

Classic Pipeline



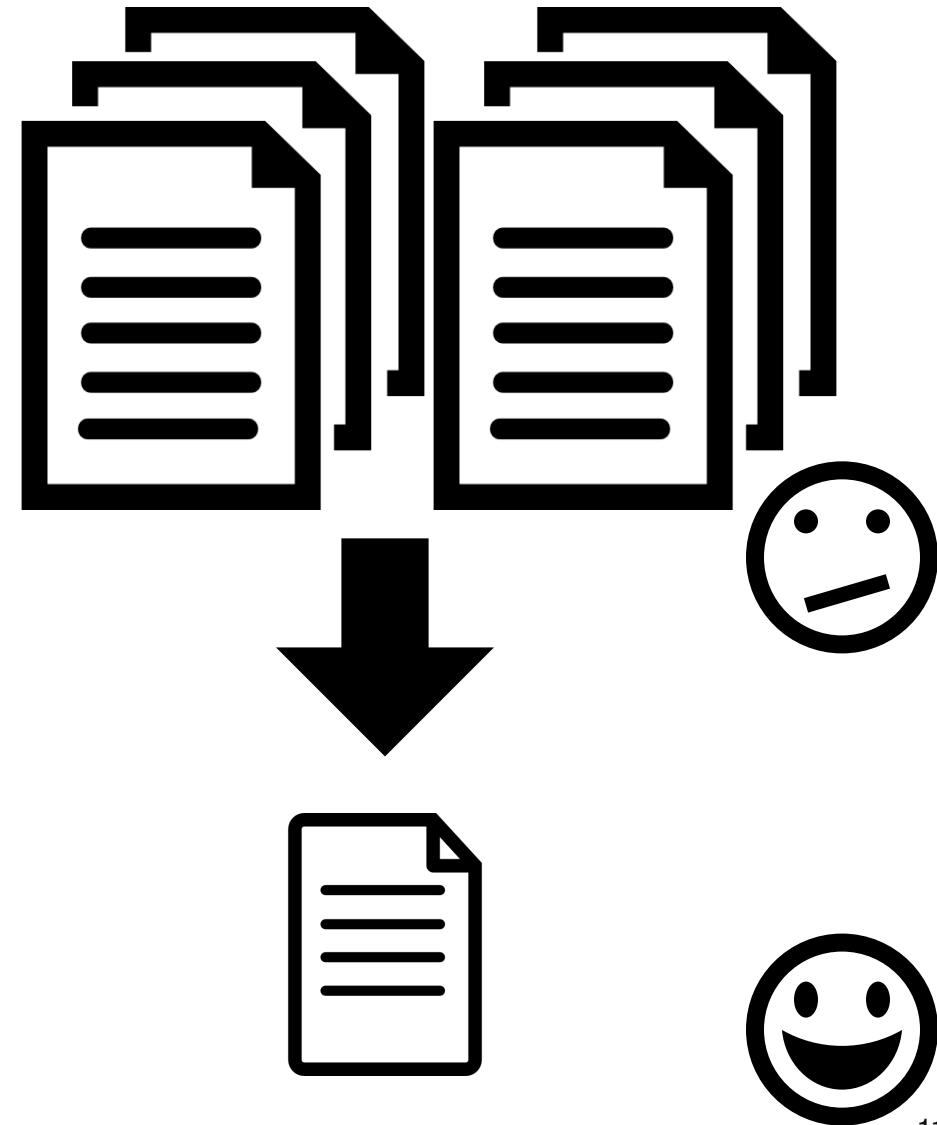
Design of each component is up to researchers or practitioners, which may involve engineering efforts.

Text Summarization



TELEGENESIS Service Excellence	
RECORDS AND FILING SUPPORT MALPRACTICE SAMPLE	
MEDICAL SUMMARY FOR PERSONAL INJURY	
Patient Name: Jane Doe	SSN: XXX-XX-XXXX
Date of Birth: 03/27/1985	Date of Injury: 06/25/2011
Current Age: 35 years	
Medical Records from mm/dd/yyyy to mm/dd/yyyy	Physicians involved in care: Dr. Tully Iskander and Dr. George Shaw (OB-GYN)
Synopsis of the case: On February 27, 2005, the patient was admitted for labor for her second child that had to be emergently converted to a caesarean section due to fetal distress. Following this surgery, she started to complain of unbearable pain in her suprapubic region which would not abate regardless of all modalities tried. On March 20, 2005, it was decided to take her in for a CT scan and it was discovered that she had a retractor forceps left in. She was immediately taken into surgery and the foreign object was taken out.	
Patient's Past Medical History: The patient was a healthy young woman G2P2A0, with no real medical problems. She had a normal spontaneous delivery with her first daughter who is now 2 years old. There were no complications at that time. She has not undergone any surgeries and not suffered any chronic illness.	
Chronological Narrative	
Date and Time	Condition and Treatment
February 27, 2005, 1100 hours	The patient started to feel contractions going into labor. She arrived at the New London Hospital at 1 pm. After initial prep the patient was taken to the Gynecology Room. On ultrasound the fetal heart rate was faster at almost 200 bpm. Decision was made to do caesarean section due to fetal distress.
February 27, 2005, 1320 hours	The patient was taken to the operating room and a healthy male was delivered with APGAR score of 10/10, weighing 7 pounds. The patient was transferred to the medical floor at 1500 hours without any complications.
February 27, 2005, 1800 hours	At 1800 hours the patient began complaining of pain in the operative area and she was administered pain killers stating that it was postoperative pain and normal. The pain subsided but did not really go away.
March 7, 2005, 1300 hours	

Text Summarization



Text Summarization



(From left) SUTD founding chairman Philip Ng, SUTD chairman Lee Tzu Yang, Minister for Education Ong Ye Kung, SUTD president, Professor Chong Tow Chong, and SUTD president emeritus, Professor Thomas Magnanti, launching the commemoration "Tree" which was wheeled in by the robot Momo. ST PHOTO: TIMOTHY DAVID

PUBLISHED JUL 10, 2019, 10:43 PM SGT



Jeanell Kiew

SINGAPORE - The Singapore University of Technology and Design (SUTD) is developing a new bachelor degree in design and artificial intelligence (AI).

The announcement was made by Education Minister Ong Ye Kung on Wednesday (July 10) at SUTD'S 10th anniversary celebration.



STcomSingapore @STcomSingapore · Jul 10
SUTD to develop new **design** and **AI** degree

SUTD to develop new design and AI degree, Singapore News & Top Singapore News -SINGAPORE - The Singapore University of Technology and Design (SUTD) is developing a new bachelor degree in design and... straitstimes.com

Text Summarization

Single-document
Summarization

Multi-document
Summarization

Text Summarization

Single-document

Extractive
Summarization

Abstractive
Summarization

Multi-document

Summarization

Text Summarization

**Extractive
Summarization**

**Abstractive
Summarization**

Summarize by selecting the most meaningful words from the documents

Text Summarization

Extractive
Summarization

Abstractive
Summarization

Summarize by paraphrasing the
important information from the documents

Other Generation Tasks

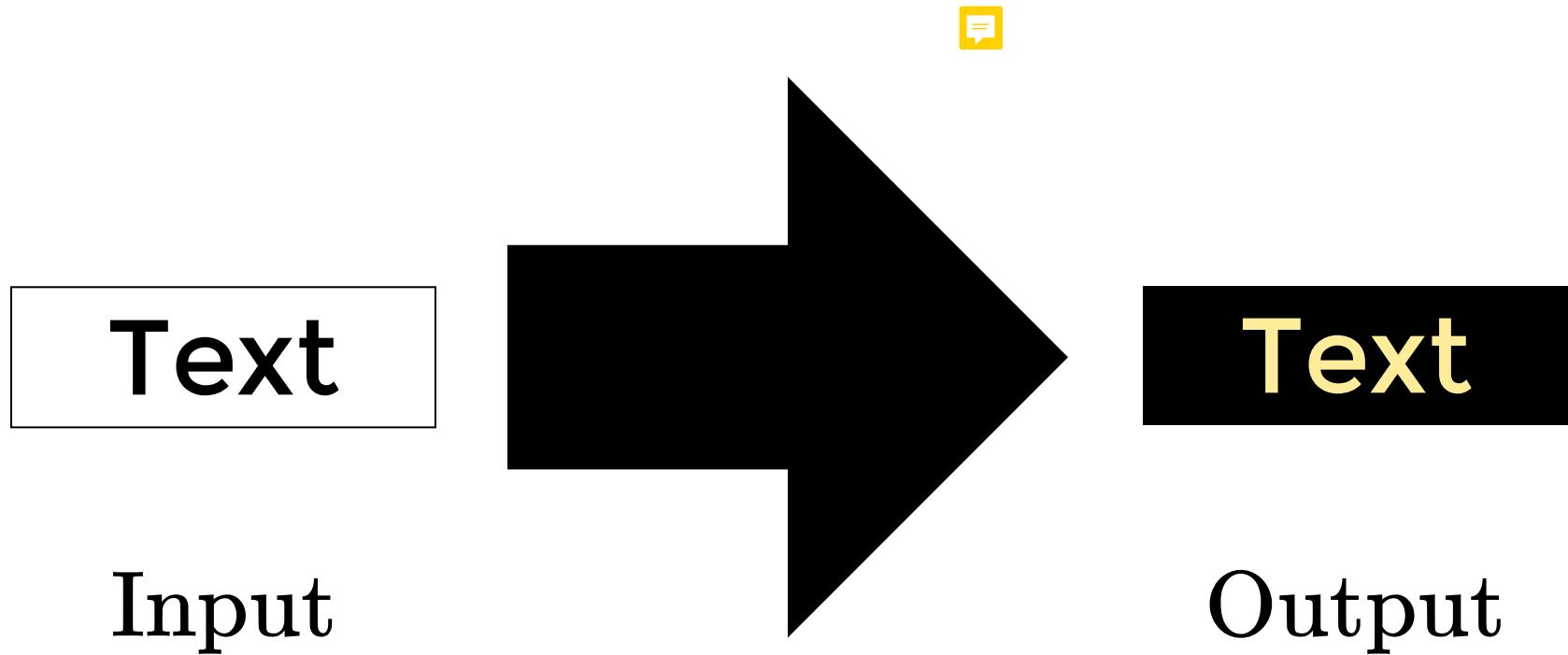
Dialogue
Systems

Question
Answering

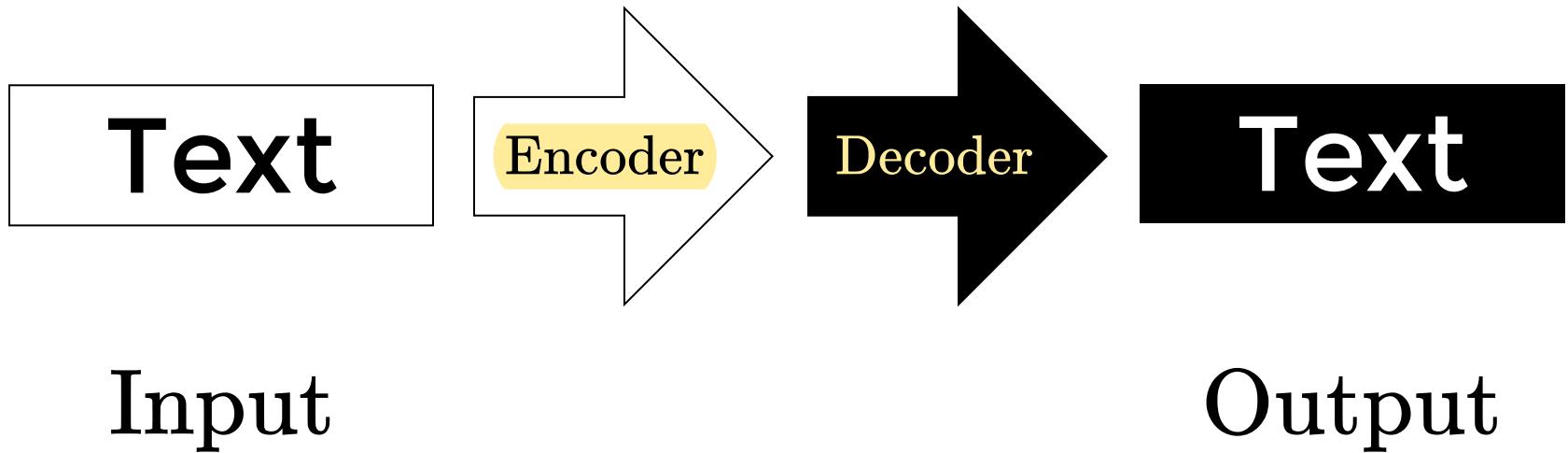
Machine
Translation

They are all text-to-text problems!

Text-to-Text Generation

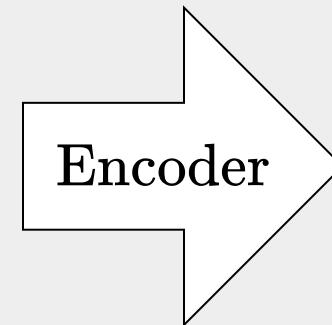
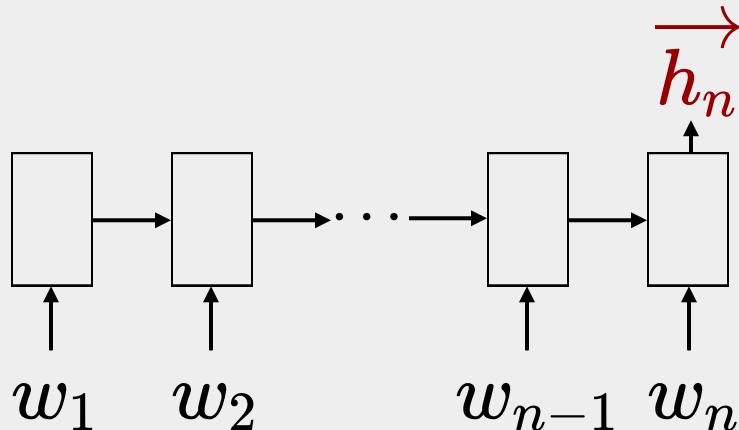


Encoder-Decoder

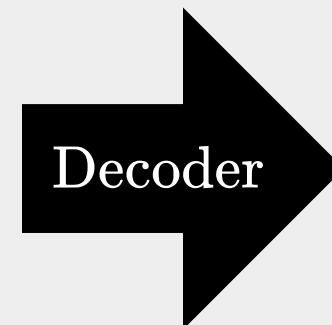
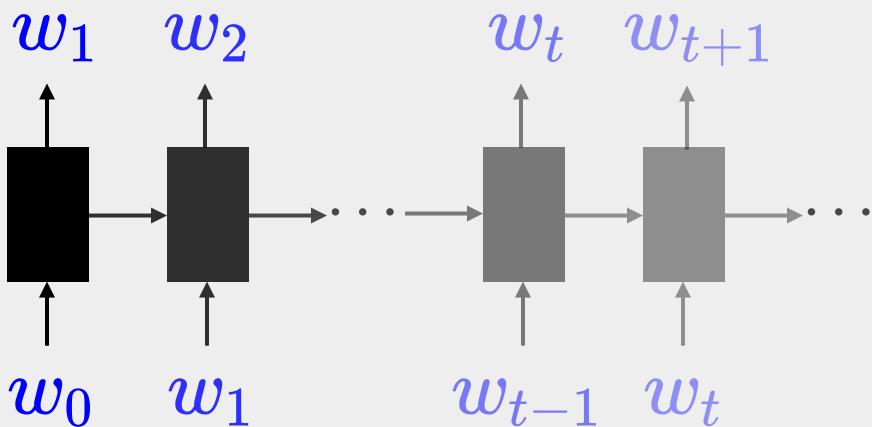


LSTM

Two applications of LSTM

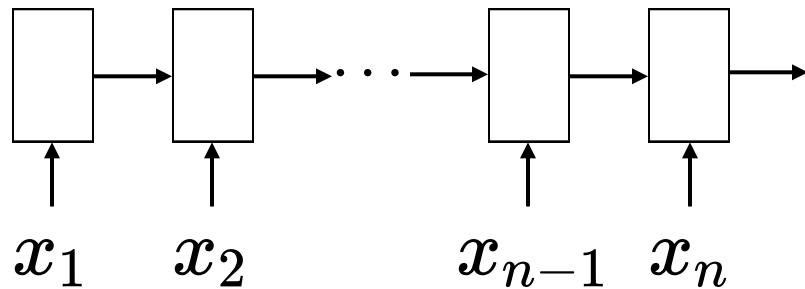


Sentence Embedding



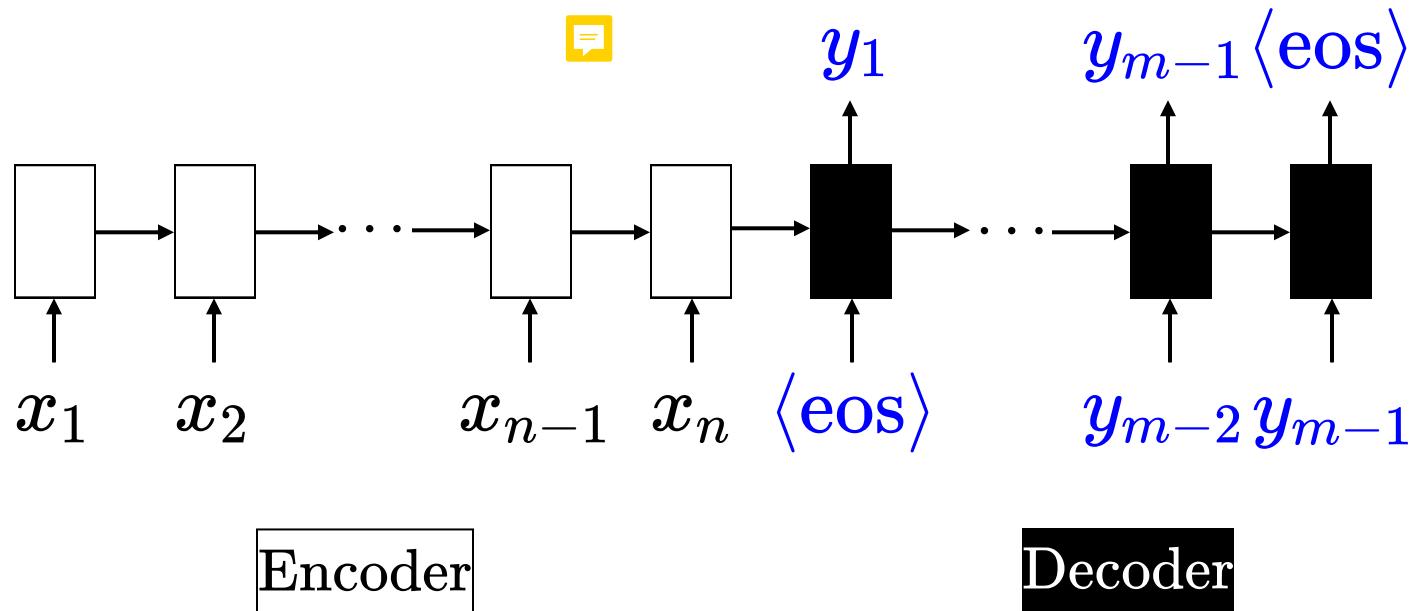
Sentence Generation

Sequence to Sequence (Sutskever et al. 2014)



Encoder

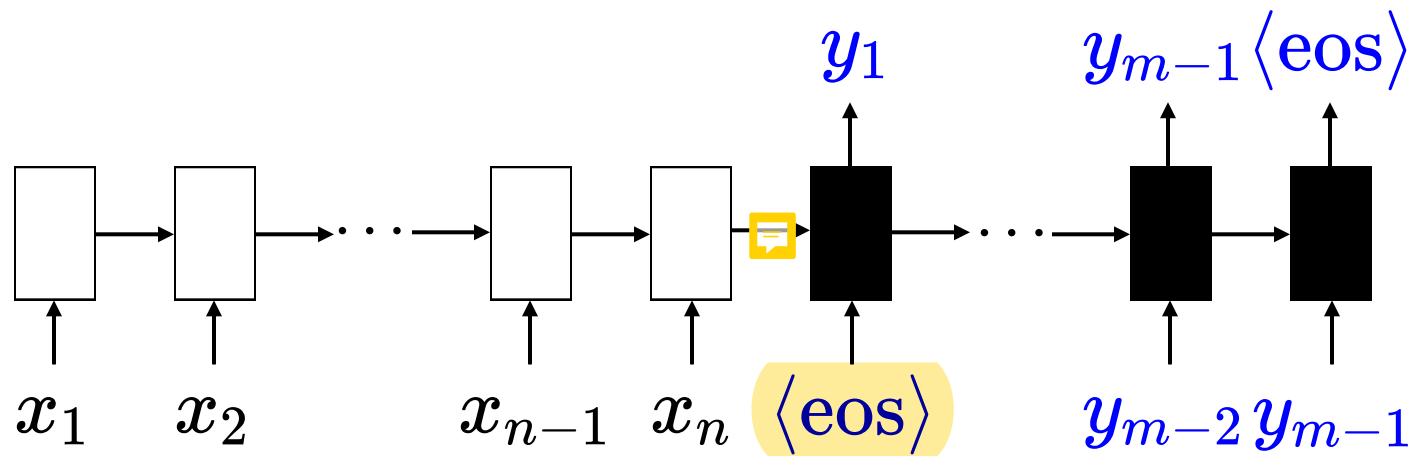
Sequence to Sequence (Sutskever et al. 2014)



Encoder

Decoder

Sequence to Sequence

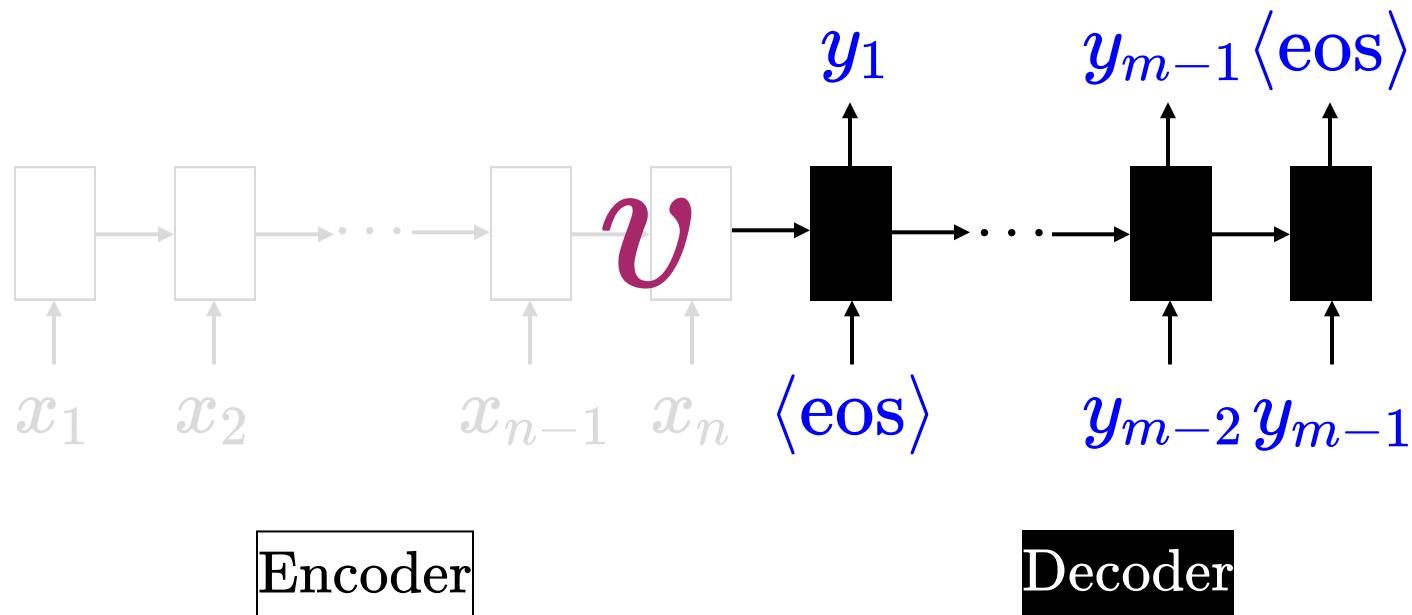


Encoder

Decoder

$$p(y_1, \dots, y_m | x_1, \dots, x_n)$$

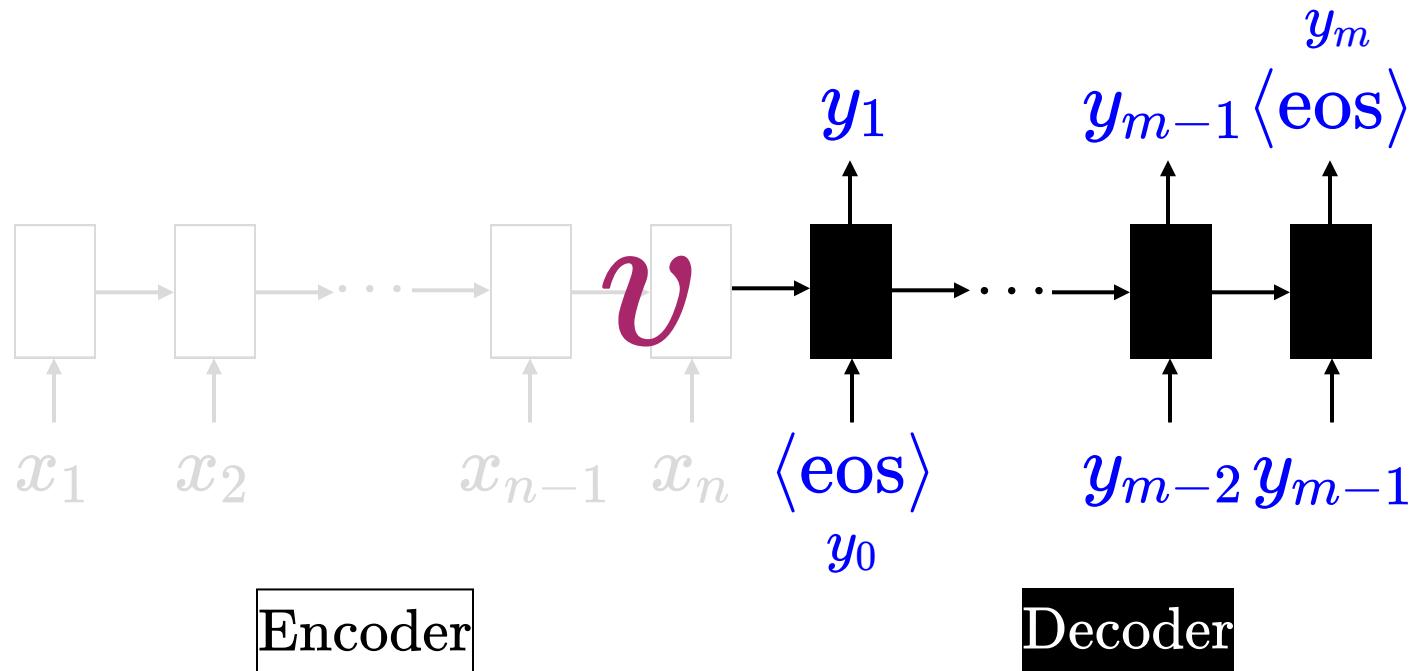
Sequence to Sequence



$$p(y_1, \dots, y_m | x_1, \dots, \mathcal{V}, x_n)$$

Sequence to Sequence

The two LSTMs will have two different sets of parameters.



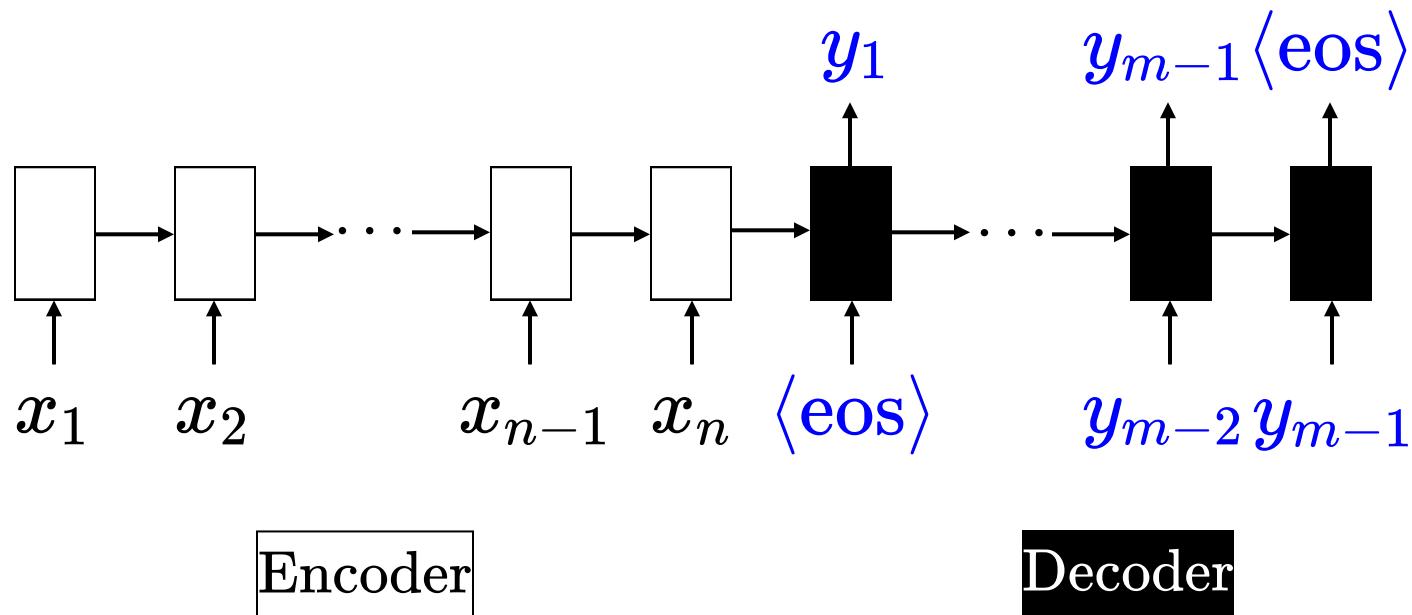
$$\prod_{k=1}^m p(y_k | v, y_0, y_1, \dots, y_{k-1})$$

Training is done with backpropagation.

Question

How do we decode based on a
new input?

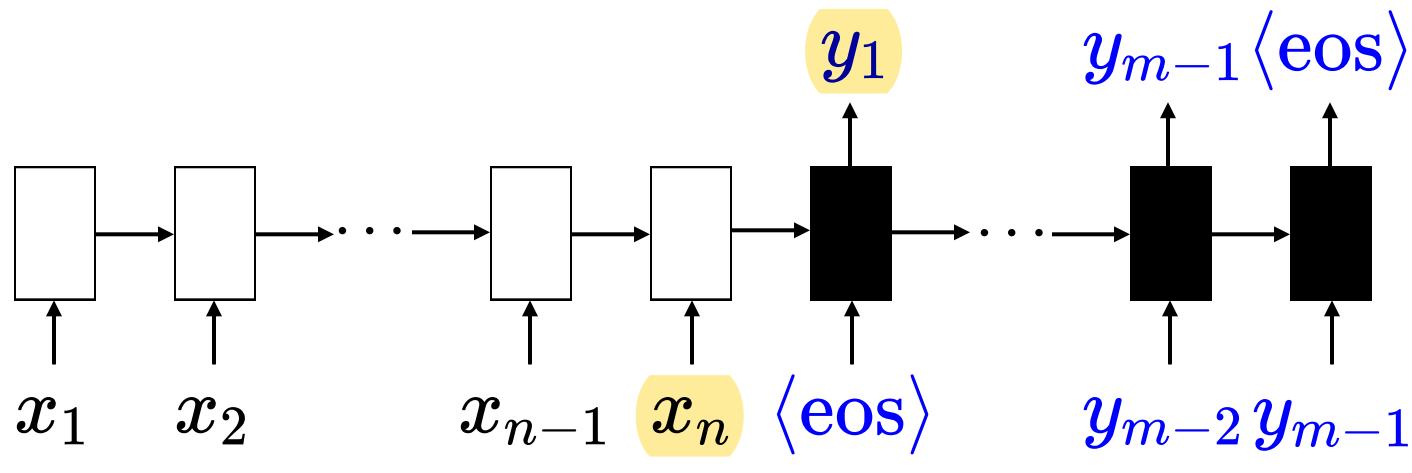
Sequence to Sequence



During decoding, we simply take a greedy approach when generating the output sequence, using the generated word as the input to the next LSTM cell.

Can we do better?

Sequence to Sequence



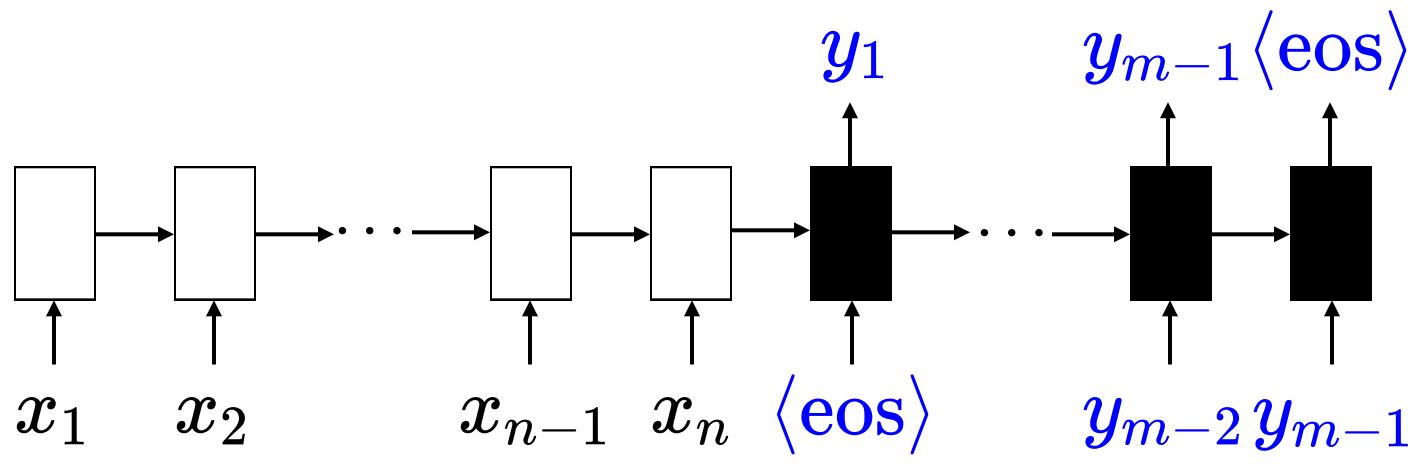
Encoder

Decoder



It was found that reversing the input sequence was very helpful. Why is that so? Can we do something better?

Sequence to Sequence



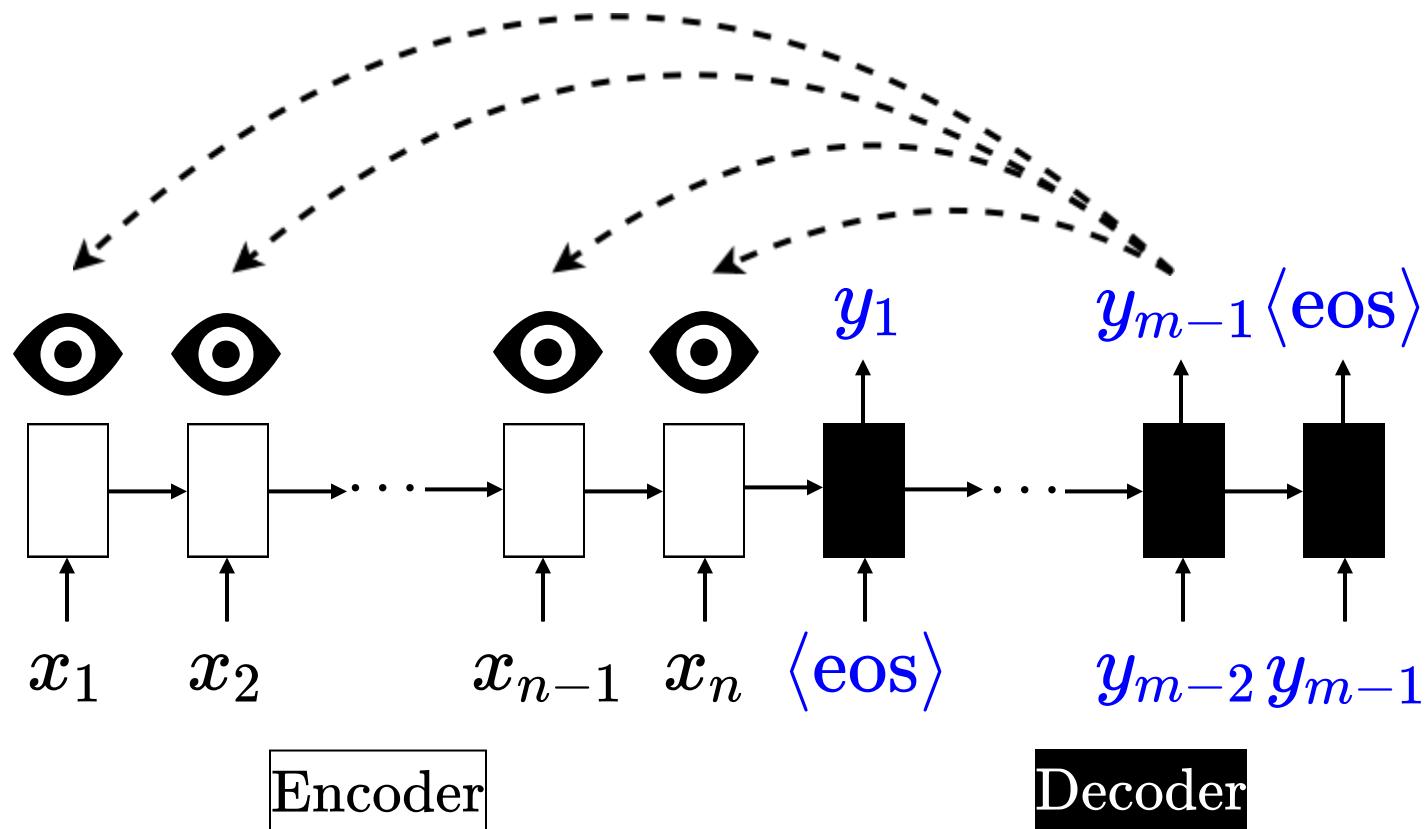
Encoder

Decoder



We may change the encoder into a bi-LSTM!
However, we need a better solution...

Seq2Seq with Attention (Bahdanau et al. 2015)

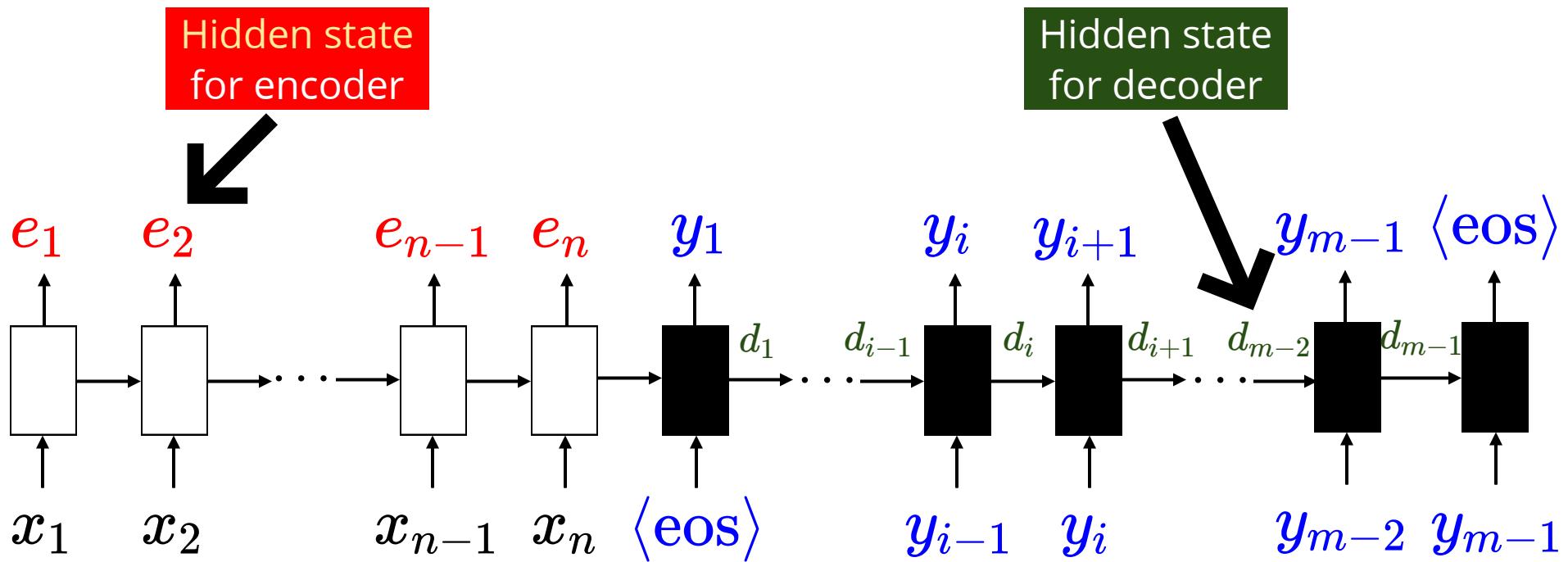


We may need to look at the input sequence
when generating the outputs...

Seq2Seq with Attention

$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$

Learnable parameters



Encoder

Decoder

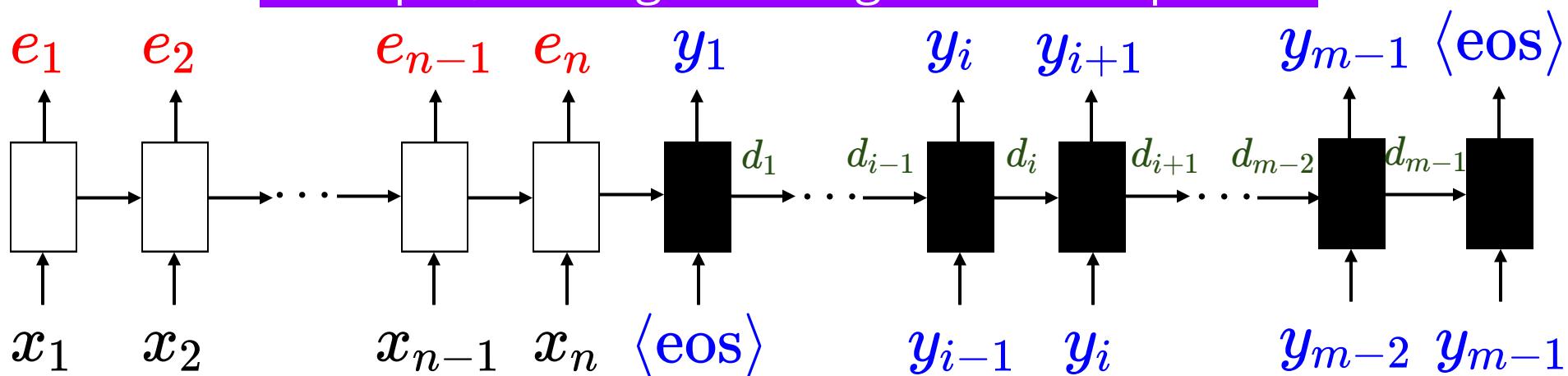
Seq2Seq with Attention

$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$

$$a_j^i = \text{softmax}(u_j^i)$$



How much I would like to focus on the j-th input, when generating the i-th output



Encoder

Decoder

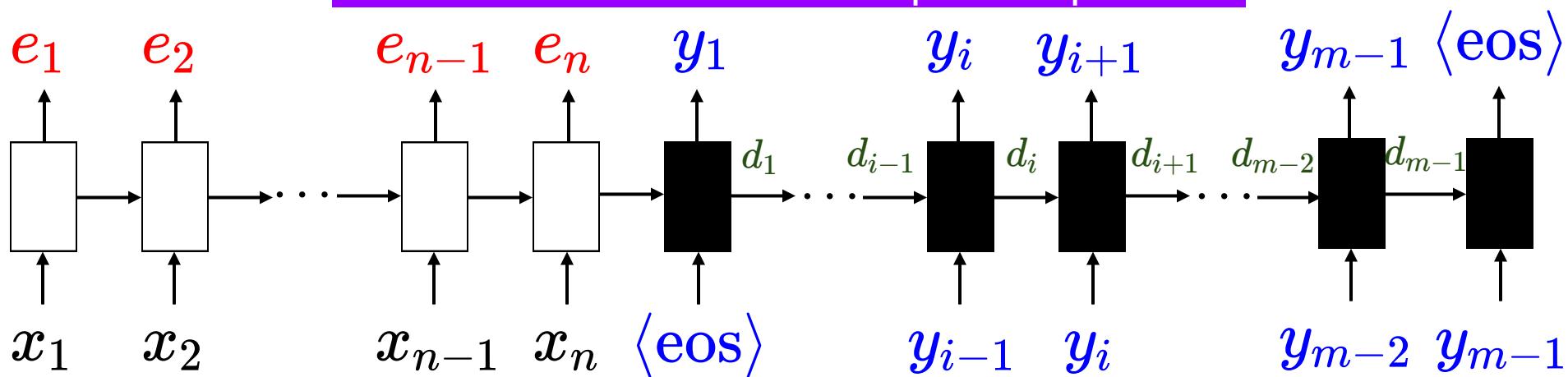
Seq2Seq with Attention

$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$

$$a_j^i = \text{softmax}(u_j^i)$$

$$d'_i = \sum_{j=1}^n a_j^i e_j$$

The weighted combination of hidden states from the entire input sequence



Encoder

Decoder

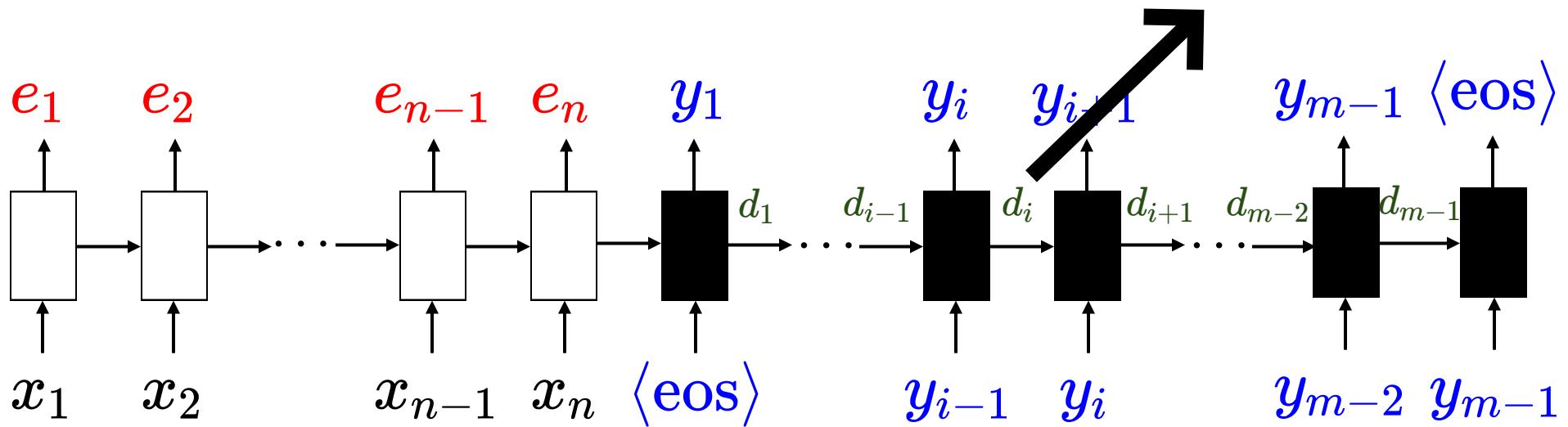
Seq2Seq with Attention

$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$

$$a_j^i = \text{softmax}(u_j^i)$$

$$d'_i = \sum_{j=1}^n a_j^i e_j$$

When predicting y_i
use $[d_i, d'_i]$ instead



Encoder

Decoder

Seq2Seq with Attention

(Vinyals and Le 2015)

Dialog System

Describe your problem: i am having issues accessing vpn

Machine: hi

Human: hello

Machine: could you please let me know what are the operating systems you are using ?

Human: i am using linux

Machine: ok it was in the ssl vpn on which you stated it was broken

Machine: please check if you are logged in to corp network

Machine: what is the error that you are running please

Human: i am seeing an error related to vpn

Machine: what is the error message that you are getting when connecting to vpn using network connect ?

Human: connection refused or something like that

Machine: may i know the version of network connect you connect ?

Human: i am not sure i know that

Machine: is the network connect prompting for any user certificate

Machine: ?

Human: yes

Machine: are you at home or at work at the moment ?

Machine: office ?

Human: i am at home

Machine: try this

Machine: goto <URL> page and enter your name and password

Human: done

Machine: try logging in with <NAME> and check if its working fine now

Human: yes , now it works !

Machine: great

Machine: anything else that i can help

Human: no , that was very useful

Machine: you are other than <NAME> :)

Machine: have a nice night

Seq2Seq with Attention (Shang et al. 2015)

Dialog System

	Post	NRM-glo	NRM-loc	NRM-hyb	Rtr.-based
1	每年元旦都来高烧。 High fever attacks me every New Year's Day	多休息，增强免疫力！ Rest more to boost your immune system.	每年元旦都去医院看病 Go to see a doctor every New Year's Day	祝早日康复，身体健康！ Get well soon and stay healthy!	高烧... High fever...
2	又老了一岁，感谢我们组，我很高兴 I gain one more year. Grateful to my group, so happy	生日快乐！永远支持你！ Happy birthday! Will stand by you forever!	生日快乐！永远年轻漂亮！ Happy birthday! Wish you always stay young and pretty!	老了老了，岁月不饶人啊！ Getting old now. Time has no mercy.	我只是老了一岁 I just gain only one more year.
3	吃素第一天，坚持住，崔朵拉。 First day of being a vegetarian. Hold on, CuiDuoLa	必须来支持！加油！ I come to offer my support. Keep it up!	坚持住，坚持下去。 Hold on, keep it up.	每天都是素食主义者 I have been being a vegetarian everyday	坚持住！ Hold on!
4	这种天气真应该带个帅锅出去踏青赏花 We should go out with some cute guys to enjoy a great outing in such a nice weather.	这是哪儿啊？真美！ Where is it? It is so beautiful!	阳光明媚，心情舒畅。 Such a nice sunny day! I am in a great mood.	这种天气真不是盖的。 It is indeed a very nice weather.	文山啊出去踏青寻找灵感哈哈 WenShan , let's go out to get some inspiration. Ha! Ha!

Input

Outputs

Simple

Seq2Seq with Attention

(Nallapati et al. 2016)

Summarization

Good quality summary output

S: a man charged with the murder last year of a british backpacker confessed to the slaying on the night he was charged with her killing , according to police evidence presented at a court hearing tuesday . ian douglas previte , ## , is charged with murdering caroline stuttle , ## , of yorkshire , england

T: man charged with british backpacker 's death confessed to crime police officer claims

O: man charged with murdering british backpacker confessed to murder

S: following are the leading scorers in the english premier league after saturday 's matches : ## - alan shearer -lrb-newcastle united -rrb- , james beattie .

T: leading scorers in english premier league

O: english premier league leading scorers

S: volume of transactions at the nigerian stock exchange has continued its decline since last week , a nse official said thursday . the latest statistics showed that a total of ##.### million shares valued at ###.### million naira -lrb- about #.### million us dollars -rrb- were traded on wednesday in , deals .

T: transactions dip at nigerian stock exchange

O: transactions at nigerian stock exchange down

Poor quality summary output

S: broccoli and broccoli sprouts contain a chemical that kills the bacteria responsible for most stomach cancer , say researchers , confirming the dietary advice that moms have been handing out for years . in laboratory tests the chemical , <unk> , killed helicobacter pylori , a bacteria that causes stomach ulcers and often fatal stomach cancers .

T: for release at #### <unk> mom was right broccoli is good for you say cancer researchers

O: broccoli sprouts contain deadly bacteria

S: norway delivered a diplomatic protest to russia on monday after three norwegian fisheries research expeditions were barred from russian waters . the norwegian research ships were to continue an annual program of charting fish resources shared by the two countries in the barents sea region .

T: norway protests russia barring fisheries research ships

O: norway grants diplomatic protest to russia

S: j.p. morgan chase 's ability to recover from a slew of recent losses rests largely in the hands of two men , who are both looking to restore tarnished reputations and may be considered for the top job someday . geoffrey <unk> , now the co-head of j.p. morgan 's investment bank , left goldman , sachs & co. more than a decade ago after executives say he lost out in a bid to lead that firm .

T: # executives to lead j.p. morgan chase on road to recovery

O: j.p. morgan chase may be considered for top job

T: target summary; O: system output

Text Summarization

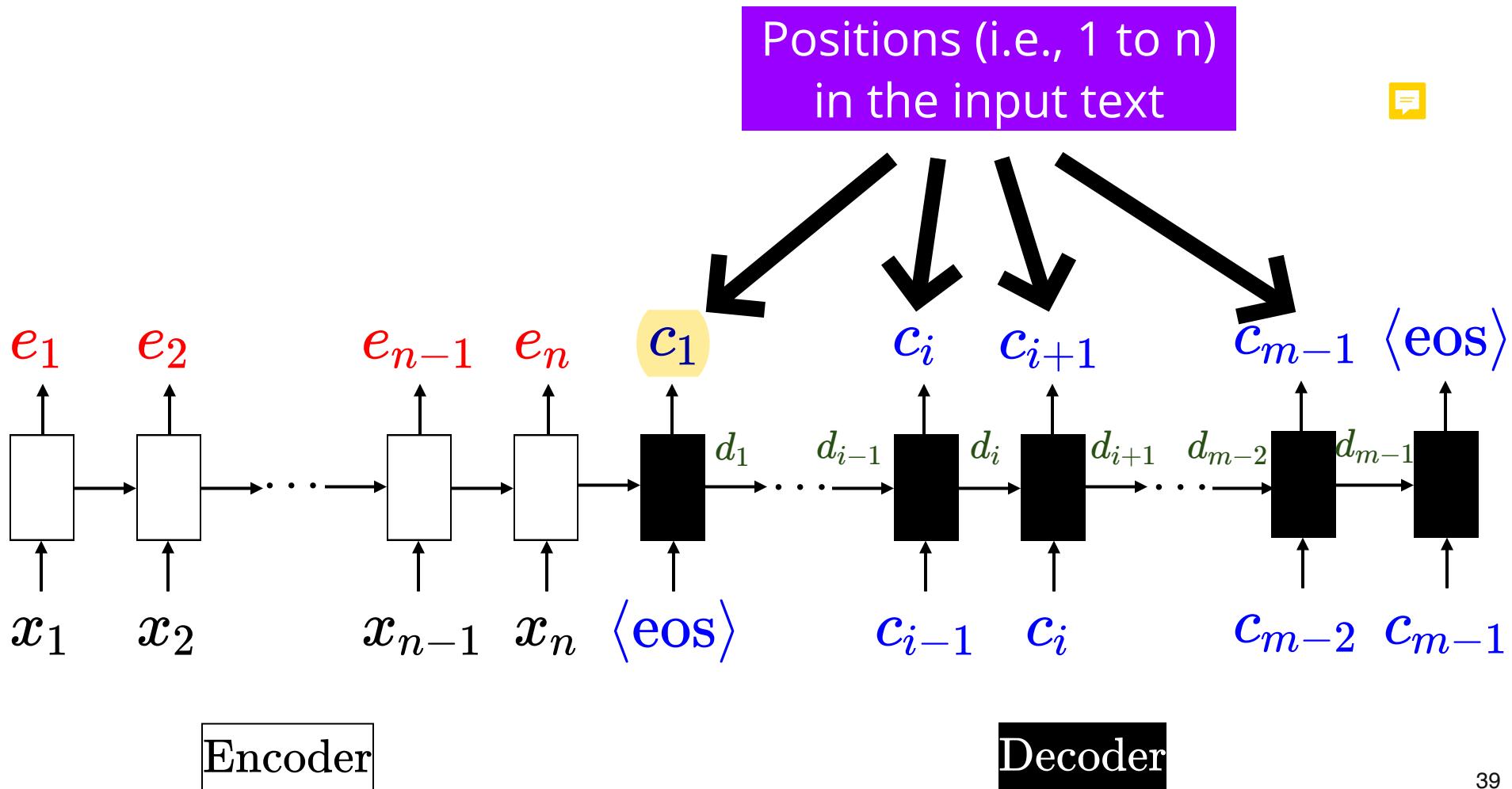
One observation
Very often, the target sentence
consists of phrases extracted from
the inputs.



Can we design an extractive summarization system
based on seq2seq?

Pointer Network

(Vinyals et al. 2015)



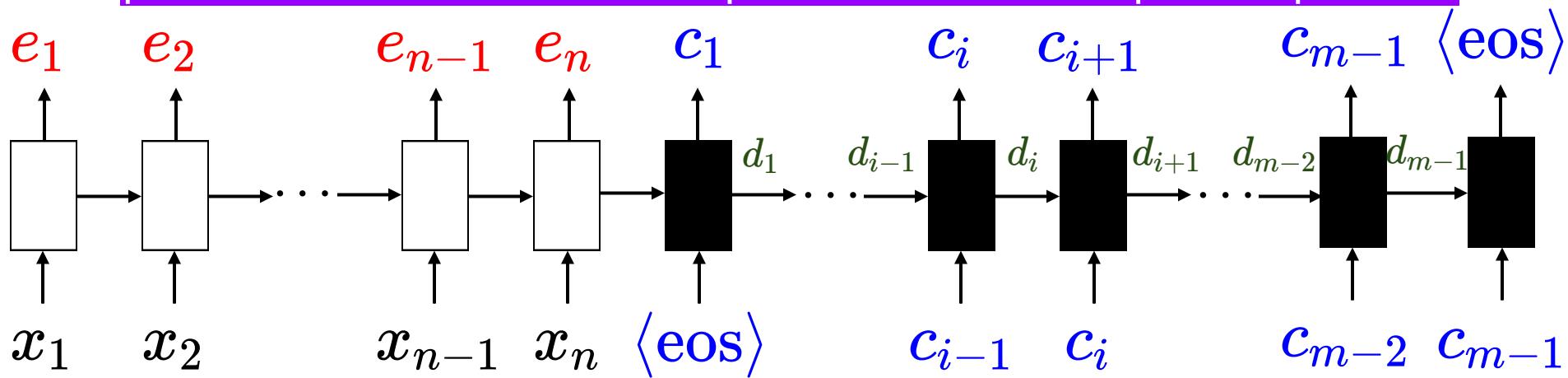
Pointer Network

(Vinyals et al. 2015)

$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$

$$p(c_i = j | c_0, \dots, c_{i-1}, x_1, \dots, x_n) = \text{softmax}(u_j^i)$$

We can extract phrases from the input by generating the positions of start and end positions in the input sequence.



Encoder

Decoder

Optional

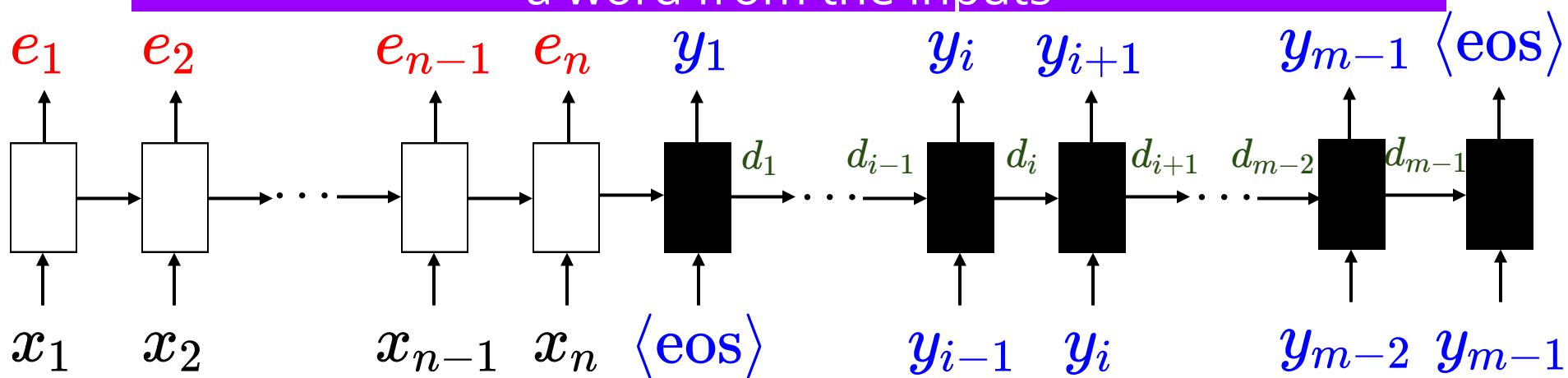
Pointer-Generator Net (See et al. 2017)

$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$

$$a_j^i = \text{softmax}(u_j^i)$$

$$p_{\text{gen}} = \sigma(W_a a^i + W_h d_i + W_y y_i + b)$$

The probability of generating a word, rather than copying a word from the inputs

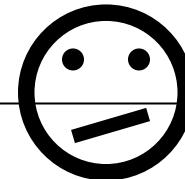


Encoder

Decoder

Evaluation

One common problem with non-structured prediction problems is there is a lack of single ground truth. Therefore, we may need human evaluation.



Evaluation

Automatic evaluations are largely based on some heuristically defined metrics.

An evaluation metric is regarded as a well-defined metric if the scores returned by the metric correlate with human judgements well.

Evaluation

ROUGE: A Package for Automatic Evaluation of Summaries

Chin-Yew Lin
Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292
cyl@isi.edu

Abstract

ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It includes measures to automatically determine the quality of a summary by comparing it to other (ideal) summaries created by humans. The measures count the number of overlapping units such as n-gram, word sequences, and word pairs between the computer-generated summary to be evaluated and the ideal summaries created by humans. This paper introduces four different ROUGE measures: ROUGE-N, ROUGE-L, ROUGE-W, and ROUGE-S included in the ROUGE summarization evaluation package and their evaluations. Three of them have been used in the Document Understanding Conference (DUC) 2001, 2002, and 2003 data. Section 7 concludes this paper and discusses future directions.

1 Introduction

Traditionally evaluation of summarization involves human judgments of different quality metrics, for example, coherence, conciseness, grammaticality, readability, and content (Mani, 2001). However, even simple manual evaluation of summaries on a large scale over a few linguistic quality questions and content coverage as in the Document Understanding Conference (DUC) (Over and Yen, 2003) would require over 3,000 hours of human efforts. This is very expensive and difficult to conduct in a frequent basis. Therefore, how to evaluate summaries automatically has drawn a lot of attention in the summarization research community in recent years. For example, Saggion et al. (2002) proposed three content-based evaluation methods that measure similarity between summaries. These methods are: cosine similarity, unit overlap (i.e. unigram or bigram), and longest common subsequence. However, they did not show how the results of these automatic evaluation methods correlate to human judgments. Following the successful application of automatic evaluation methods, such as BLEU (Papineni et al., 2001), in machine translation evaluation, Lin and Hovy (2003) showed that methods similar to BLEU,

i.e. n-gram co-occurrence statistics, can be used to evaluate summaries. In this paper, we present a package, ROUGE, for automatic evaluation of summaries and its evaluations. ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It includes several automatic evaluation measures to measure the similarity between summaries. We describe ROUGE-N in Section 2, ROUGE-L in Section 3, ROUGE-W in Section 4, and ROUGE-S in Section 5. Section 6 shows how these measures correlate with human judgments using DUC 2001, 2002, and 2003 data. Section 7 concludes this paper and discusses future directions.

2 ROUGE-N: N-gram Co-Occurrence Statistics

Formally, ROUGE-N is an n-gram recall between a candidate summary and a set of reference summaries. ROUGE-N is computed as follows:

$$\text{ROUGE-N} = \frac{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{gram}_n \in S} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{gram}_n \in S} \text{Count}(\text{gram}_n)} \quad (1)$$

Where n stands for the length of the n-gram, gram_n , and $\text{Count}_{\text{match}}(\text{gram}_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries.

It is clear that ROUGE-N is a recall-related measure because the denominator of the equation is the total sum of the number of n-grams occurring at the reference summary side. A closely related measure, BLEU, used in automatic evaluation of machine translation, is a precision-based measure. BLEU measures how well a candidate translation matches a set of reference translations by counting the percentage of n-grams in the candidate translation overlapping with the references. Please see Papineni et al. (2001) for details about BLEU.

Note that the number of n-grams in the denominator of the ROUGE-N formula increases as we add more references. This is intuitive and reasonable because there might exist multiple good summaries.



ROUGE: A widely adopted evaluation metric for text summarization

Evaluation



ROUGE-N: Overlap of N-grams
between the system and
reference summaries.

ROUGE-L: Overlaps based on
longest common subsequence.
More flexible measure.

ROUGE-S: Overlap of skip-
bigrams (two words that appear
together in sentence order).

Tasks in NLP

