Applications of Recurrent Neural Networks in Natural Language Processing

Isabelle Augenstein
isabelleaugenstein.github.io
@IAugenstein

UNIVERSITY OF COPENHAGEN

RNNs/LSTMs – What can we learn?

- Sequence generation
  - One prediction (word/character) for each part of sequence
- Sequence tagging
  - One prediction (label) for each part of sequence
- Sequence classification
  - One prediction (label) for the whole sequence
- Sequence-to-sequence
  - One sequence of predictions (words/characters) given a sequence (words/characters)

RNNs/LSTMs – What can we learn?

- Sequence generation
  - One prediction (word/character) for each part of sequence
- Sequence tagging
  - One prediction (label) for each part of sequence
- Sequence classification
  - One prediction (label) for the whole sequence
- Sequence-to-sequence
  - One sequence of predictions (words/characters) given a sequence (words/characters)

Generative Learning: Language Models

- What word(s) should come next?
- *We _*
- *We are going _*
- *We are going to win _*

Generative Learning: Language Models

- Calculate the probability of seeing a sequence

We are going to win bigly

vs.

We are going to win big league

LM Use Cases: Speech Recognition

- What did he say?

We are going to win bigly

vs.

We are going to win big league

LM Use Cases: Machine Translation

- What should it be translated to?

Wir werden haushoch gewinnen

->

We are going to win bigly

vs.

We are going to win by a mile

LM Use Cases: Language Generation

LM Use Cases: Machine Translation

- What should it be translated to?

Wir werden haushoch gewinnen

->

We are going to win bigly

vs.

We are going to win by a mile

Language Models, Formally

Model the probability

$p(w_1, \ldots, w_d)$

of observing the sequence $w_1, \ldots, w_d$:

$$p(w_1, \ldots, w_d) = p(w_1)p(w_2|w_1)p(w_3|w_1,w_2)\ldots$$

$$= p(w_1) \prod_{i=2}^{d} p(w_i|w_1, \ldots, w_{i-1})$$

Structured prediction:

predict word y = w_1 *conditioned on history* $w_1, \ldots, w_{i-1}$

Language Models & RNNs

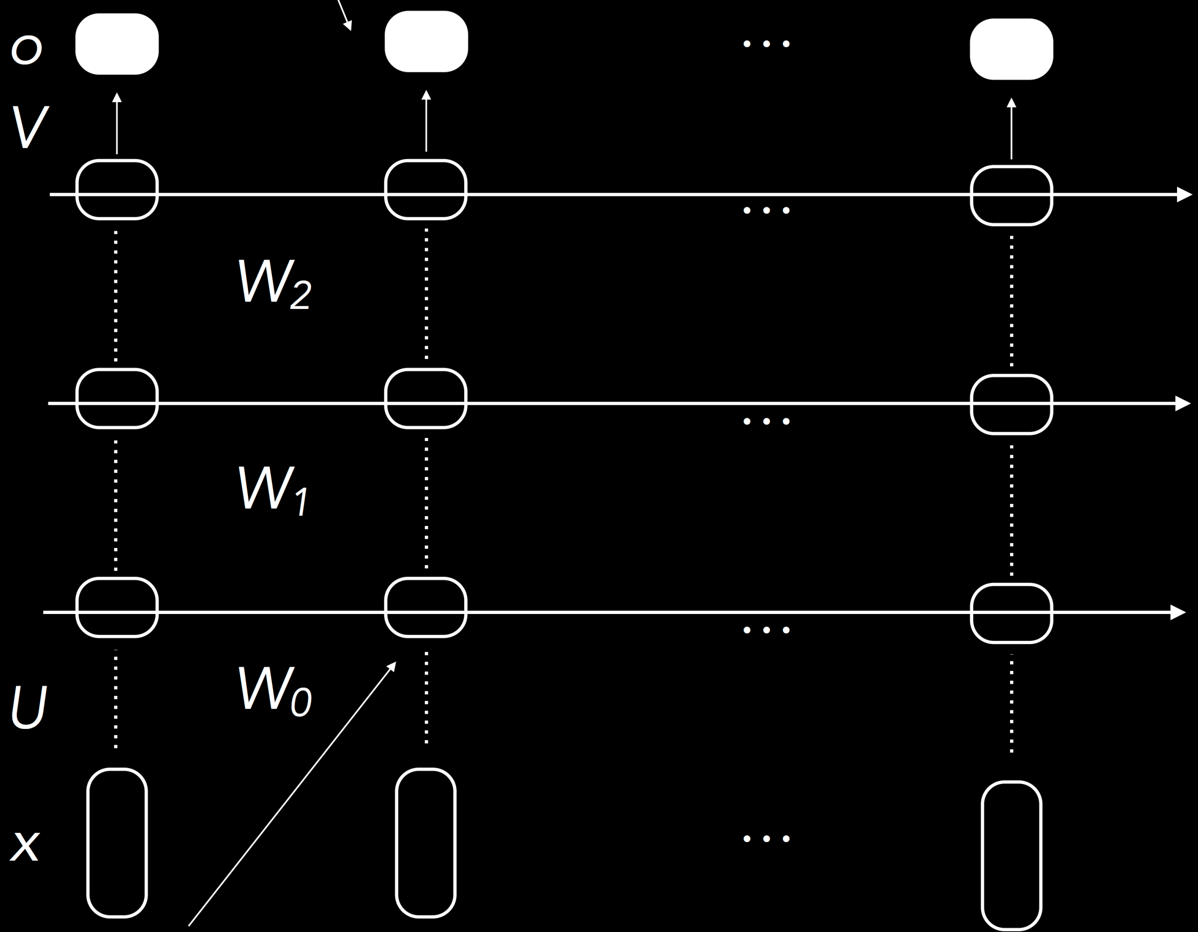
Probability

$p(w_1, \ldots, w_d)$

is obtained through softmax function

Language Models & RNNs

$$o_t = \mathrm{softmax}(V s_t)$$



$$s_t = f(U x_t + W s_{t-1})$$

Sampling from Language Models

- Sample / generate sequence incrementally, one word at a time:
- *We _*
- *We are _*
- *We are going _*

…

Evaluating Language Models

Extrinsic: how does it improve a downstream task (e.g. speech recognition, machine translation)?

Intrinsic: how well does it model language?

Intrinsic Evaluation of Language Models: Perplexity

Shannon Game: predict next word, win if prediction match words in actual corpus (or you gave it high probability)

Perplexity: for sequence $w_1, \ldots, w_T$ *of T words*

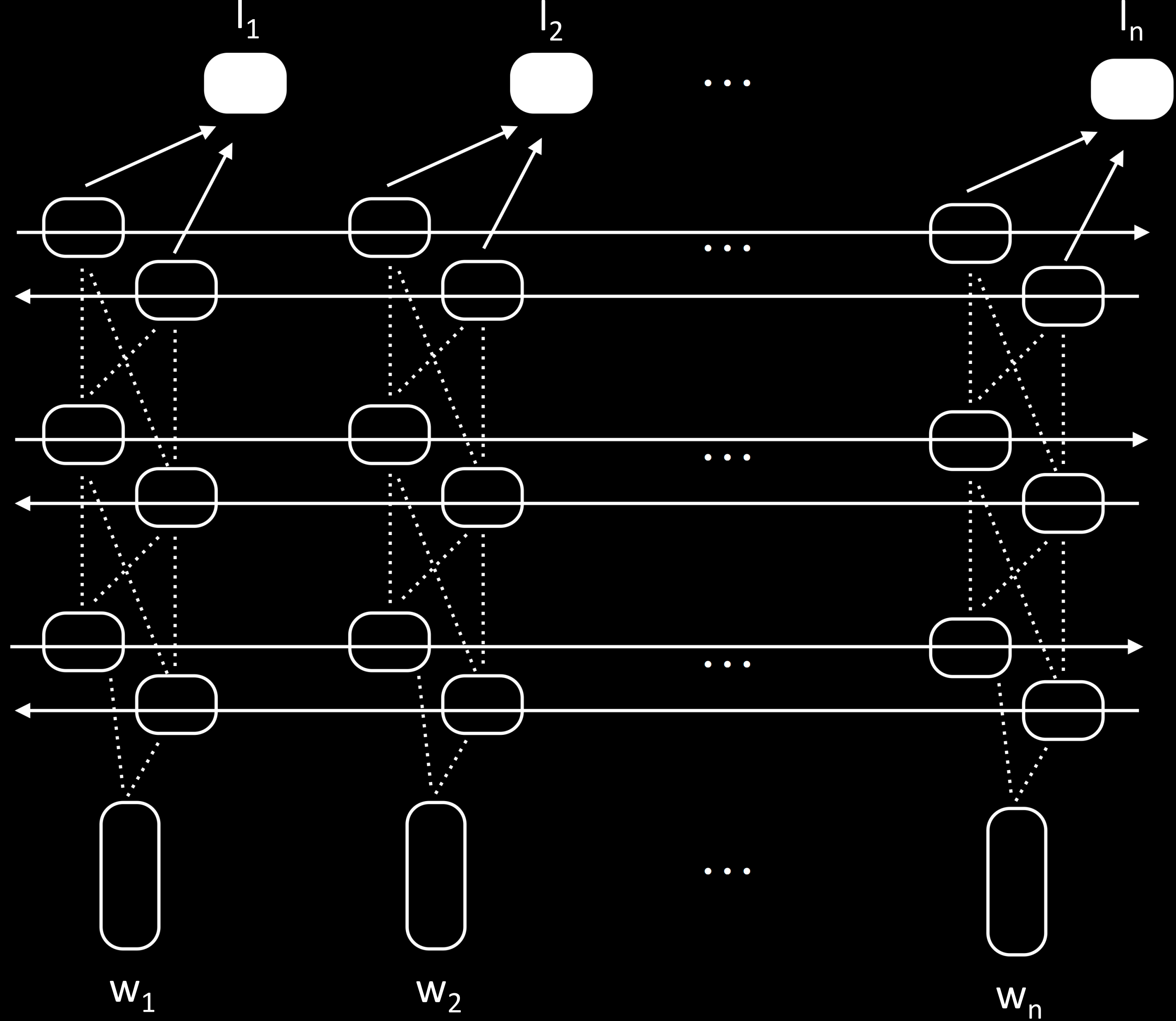
$$P(w_1,\ldots,w_T) = p(w_1, \ldots, w_T)^{-1/T} =$$

$$\sqrt[T]{\prod 1 / p(w_1 \mid w_{1-n}, \ldots, w_{i-1}))}$$

RNNs/LSTMs – What can we learn?

- Sequence generation
  - One prediction (word/character) for each part of sequence
- Sequence tagging
  - One prediction (label) for each part of sequence
- Sequence classification
  - One prediction (label) for the whole sequence
- Sequence-to-sequence
  - One sequence of predictions (words/characters) given a sequence (words/characters)

Sequence tagging



Sequence tagging

- Assign a label to each token in sequence

- Part of speech tagging
- Named entity recognition

Demo:

http://nlp.stanford.edu:8080/corenlp/process

Sequence tagging: Applications

- Text Mining: what concepts are mentioned in text?
  - Musical artists, books, … -> recommender systems
  - Locations -> recommender systems, disaster relief
  - Politicians, parties -> political analysis

Sequence tagging: Challenges

Example: Time flies like an arrow

- Time moves in a way an arrow does
- Measure the speed of flies like an arrow
- Time flies, a type of flies, like an arrow
- …

https://en.wikipedia.org/wiki/Time_flies_like_an_arrow;_fruit_flies_like_a_banana

Sequence tagging: Challenges

Example: The old man the boat

Example: The horse raced past the barn fell

Example: The complex houses married and single soldiers and their families

https://en.wikipedia.org/wiki/Garden_path_sentence

POS tags in Penn Treebank

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential there
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun

Number	Tag	Description
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	to
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

POS tags in Penn Treebank

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential there
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun

Number	Tag	Description
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	to
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Named Entity Recognition: Types

- Person
- Location (country, city, river, …)
- Organisation (company, institution, …)
- Misc (everything else)

Named Entity Recognition

- Challenge: consist of several tokens

- Solution: Tag each token as either:
  - **B** (eginning of phrase)
  - **I** (nside of phrase)
  - **O** (utside of phrase)

Named Entity Recognition

- … in addition to types

Isabelle Augenstein is an asst. prof. at DIKU .
B-PER I-PER O O O O O B-ORG O

- At test time, decode

Isabelle Augenstein is an asst. prof. at DIKU .
PER ORG

Sequence tagging & RNNs

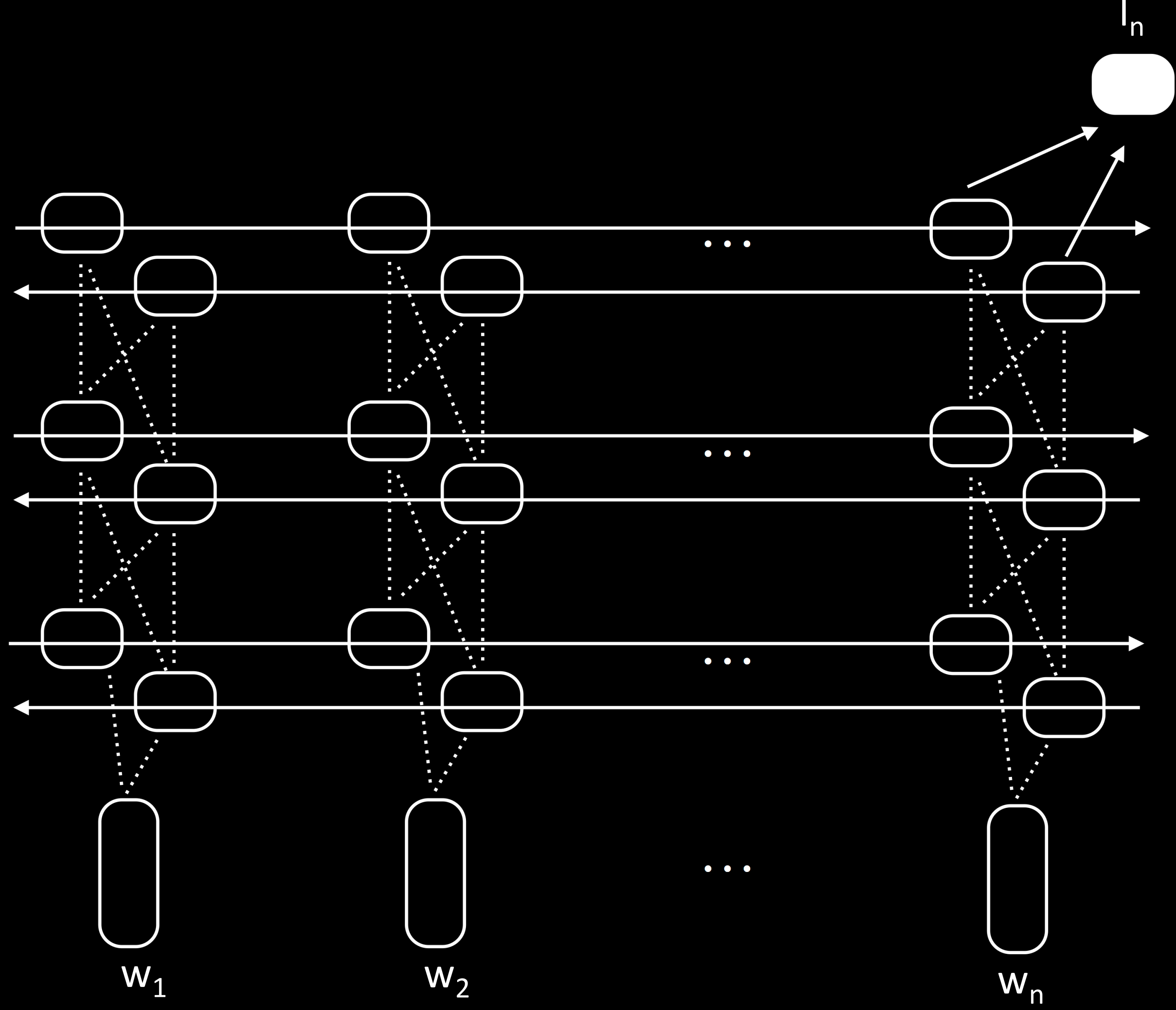
- Each state contains representation of token *in context*

- Wide context window (typically whole sentence) helps to assign correct tag to words

RNNs/LSTMs – What can we learn?

- Sequence generation
  - One prediction (word/character) for each part of sequence
- Sequence tagging
  - One prediction (label) for each part of sequence
- Sequence classification
  - One prediction (label) for the whole sequence
- Sequence-to-sequence
  - One sequence of predictions (words/characters) given a sequence (words/characters)

Sequence classification



Sequence classification

- Application: sentiment analysis

BBC News (World) @BBCWorld · 21. aug.
Kim Wall died in accident on submarine, Danish inventor says

Miki Perkins @perkinsmiki Følg

Kim Wall is not some quirky Scandi story for our consumption. She's a woman murdered while doing her job as a journalist. #vaw

Another Susan Davis @sadele2 · 32 min.
Kim Wall: An 'exceptional' journalist. Family says: Remember her work, not her murder. Links to 4 of her stories

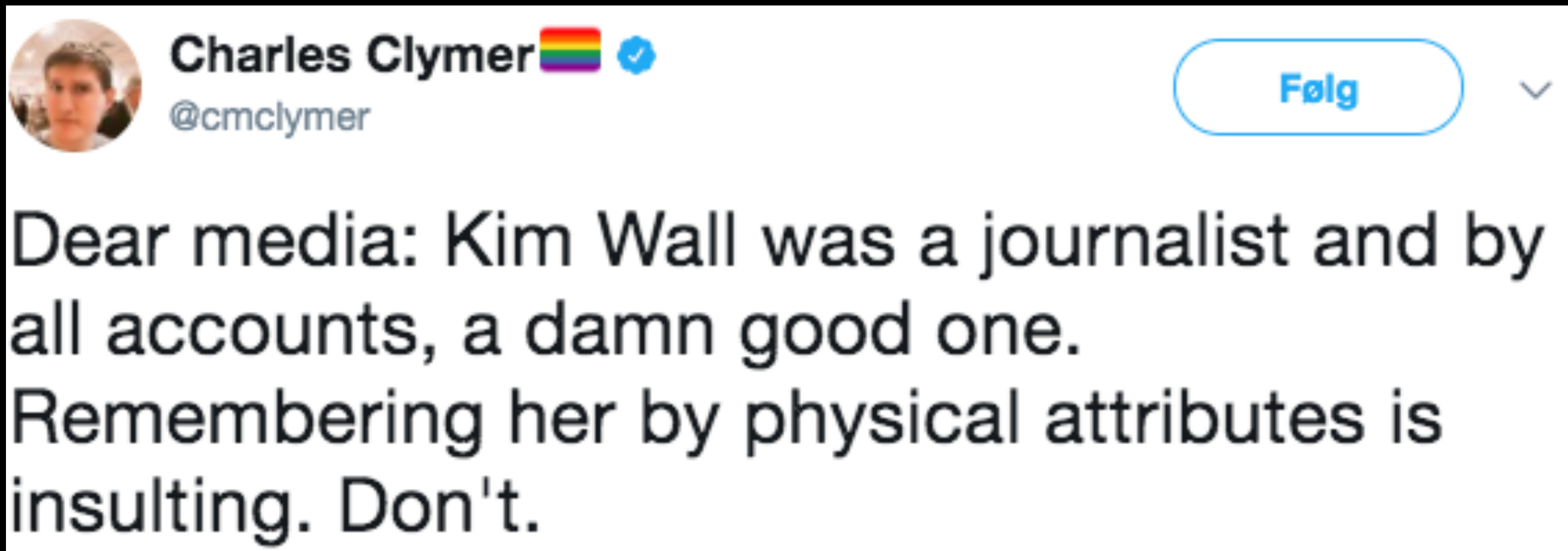
Sequence classification

- Application: topic classification

- Application: event detection

Sequence classification: challenges

- Labels change within sequence



Sequence classification: challenges

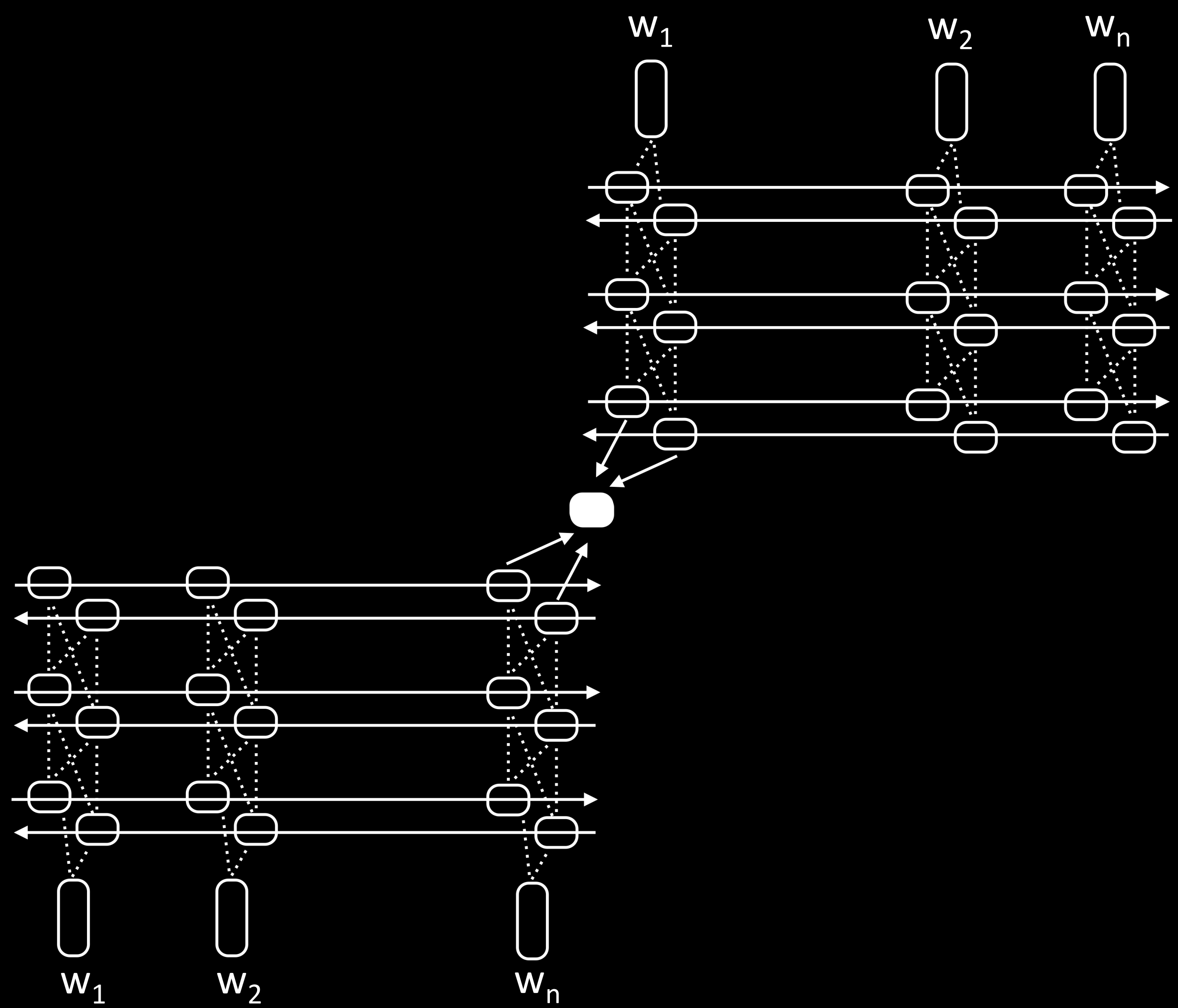
- Topic is not mentioned in sequence

Headless torso of a woman found floating in Danish waters is that of missing Swedish journalist

RNNs/LSTMs – What can we learn?

- Sequence generation
  - One prediction (word/character) for each part of sequence
- Sequence tagging
  - One prediction (label) for each part of sequence
- Sequence classification
  - One prediction (label) for the whole sequence
- Sequence-to-sequence
  - One sequence of predictions (words/characters) given a sequence (words/characters)

Sequence-to-sequence



Sequence-to-sequence: Applications

- Machine Translation

https://translate.google.com/

Sequence-to-sequence: Applications

- Text Simplification

Copenhagen (Danish: København [købm̩ˈhɑwˀn]; Latin: Hafnia) is the capital and most populous city of Denmark. The city has a population of 763,908 (as of December 2016), of whom 601,448 live in the Municipality of Copenhagen. The larger urban area has a population of 1,280,371 (as of 1 January 2016), while the Copenhagen metropolitan area has just over 2 million inhabitants. Copenhagen is situated on the eastern coast of the island of Zealand; another small portion of the city is located on Amager, and is separated from Malmö, Sweden, by the strait of Øresund. The Øresund Bridge connects the two cities by rail and road.

https://en.wikipedia.org/wiki/Copenhagen

Copenhagen is the capital city of Denmark. It is also the largest city in Denmark. In 2014, 1,246,611 people lived in the urban area. Copenhagen is on the islands of Zealand and Amager.

https://simple.wikipedia.org/wiki/Copenhagen

Sequence-to-sequence: Challenges

- Long Sequences (e.g. in document summarisation)

http://summarising-scientific-papers-examples.s3-website-eu-west-1.amazonaws.com/

Sequence-to-sequence: Challenges

- Low Resources (e.g. for machine translation)

http://www.err.ee/612882/eesistumise-valismaalased-riigilipu-lehvitamine-euroopas-on-natsionalism

Thank you!
Questions?

isabelleaugenstein.github.io
augenstein@di.ku.dk
@iaugenstein
github.com/isabelleaugenstein

Tutorial Resources

- https://tinyurl.com/ybwwv9ox or Absalon

Deep Learning in NLP Framework:

- Jack the Reader – A Machine Reading Framework, https://github.com/uclmr/jack
  - Machine Reading / Question Answering can be approach as sequence tagging, sequence classification *OR* a sequence-to-sequence model!

Tutorial Task

- Example of Sequence Classification Task: Stance Detection

Target: *legalization of abortion*
Tweet: *A foetus has rights too! Make your voice heard.*

Stance*: FAVOR / AGAINST / NONE*