IN5550: Neural Methods in Natural Language Processing
Sub-lecture 6.1
Capturing language structure

Andrey Kutuzov

University of Oslo

24 February 2022



Contents



1. Obligatory 1 results
2. Capturing the structure

Obligatory 1 results



26 students submitted Obligatory 1, out of 37 enrolled in IN5550.

Obligatory 1 results



26 students submitted Obligatory 1, out of 37 enrolled in IN5550.

Leaderboard (macro-F1 scores on the test set)

1. Tellef: 0.53
2. Anna & Elina: 0.49
3. Herman: 0.48

See the solution and the test set in the GitHub repository.

Our Roadmap



Today

- Language structure: sequences, trees, graphs
- Recurrent Neural Networks
- Learning to forget: Gated RNNs

Next Week

- RNN applications:
  - Different types of sequence labeling
  - RNNs for structured prediction

Later

- Convolutional neural networks (CNNs)
- A CNN & RNN marriage: transformer architectures
- Encoder–Decoder Models, Conditioned generation and Attention
- Contextualized embeddings and transfer learning

Contents



1 Obligatory 1 results

2 Capturing the structure

So far: MLPs + embeddings as inputs



- Embeddings have benefits over discrete feature vectors; can make use of unlabeled data + information sharing across features.
- MLPs can learn certain feature combinations...

So far: MLPs + embeddings as inputs



- Embeddings have benefits over discrete feature vectors; can make use of unlabeled data + information sharing across features.
- MLPs can learn certain feature combinations...
- ...but there are limitations when it comes to learning ordering and positional information.
- We still lack power for representing sentences and documents.
- Summing or averaging? Gives a fixed-length representation, but no information about order or structure.
- Concatenation? Would blow up the parameter space for a fully connected layer.

So far: MLPs + embeddings as inputs

- We need specialized NN architectures that extract higher-level features:
- RNNs and CNNs – the agenda for the coming weeks.
- They learn intermediate representations that are then plugged into additional layers for prediction.
- Layers and architectures are like Lego bricks that plug into each-other – mix and match. *PyTorch* makes it easy.



Example text classification tasks



Document- / sentence-level polarity; positive or negative?

- *The food was expensive but hardly impressive.*
- *The food was hardly expensive but impressive.*

Example text classification tasks



Document- / sentence-level polarity; positive or negative?

- *The food was expensive but hardly impressive.*
- *The food was hardly expensive but impressive.*

- Bag of words or embeddings will not help.
- Bag-of-*n*-grams or *n*-gram embeddings?
  - Potentially wastes many parameters; only a few n-grams relevant.
  - Data sparsity issues + does not scale to higher order *n*-grams.
- We need to take word order into consideration in a more principled and general way.

Example text classification tasks



Document- / sentence-level polarity; positive or negative?

- *The food was expensive but hardly impressive.*
- *The food was hardly expensive but impressive.*

- Bag of words or embeddings will not help.
- Bag-of-*n*-grams or *n*-gram embeddings?
  - Potentially wastes many parameters; only a few n-grams relevant.
  - Data sparsity issues + does not scale to higher order *n*-grams.
- We need to take word order into consideration in a more principled and general way.

Many text classification tasks have similar traits:. . .

- authorship attribution
- spam detection
- abusive language
- subjectivity classification
- question type detection . . .

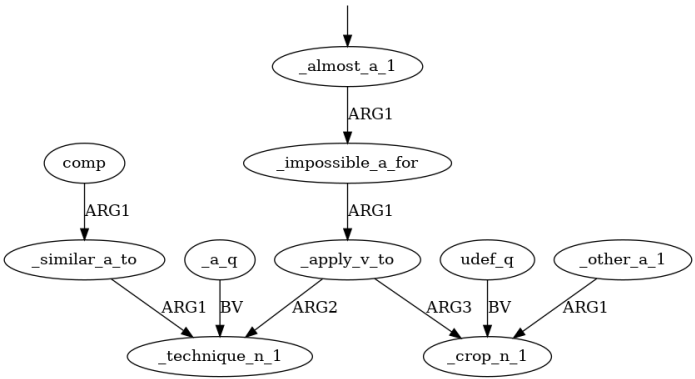
Language is So Rich in Structure



'A similar technique is almost impossible to apply to other crops.'

Language is So Rich in Structure

'A similar technique is almost impossible to apply to other crops.'



Elementary dependency structures: semantics of the sentence

http://mrp.nlpl.eu/index.php?page=2

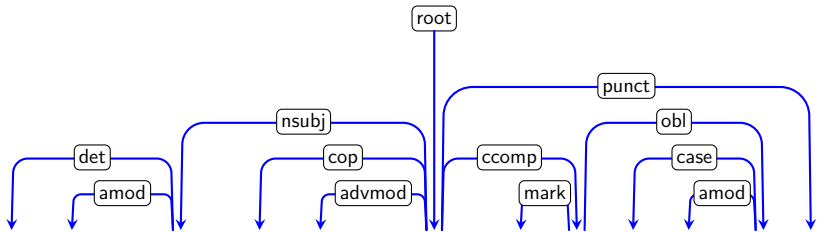
Okay, Maybe Start with Somewhat Simpler Structures



'A similar technique is almost impossible to apply to other crops.'

Okay, Maybe Start with Somewhat Simpler Structures

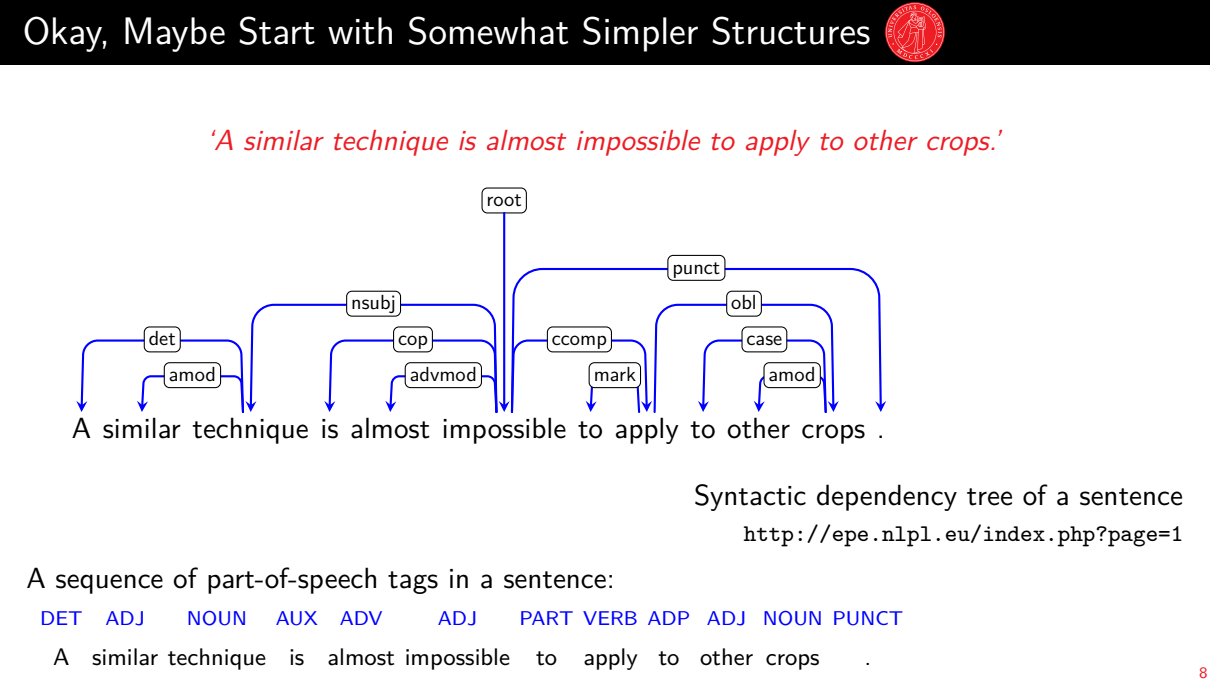
'A similar technique is almost impossible to apply to other crops.'



Syntactic dependency tree of a sentence

`http://epe.nlpl.eu/index.php?page=1`

Okay, Maybe Start with Somewhat Simpler Structures



'A similar technique is almost impossible to apply to other crops.'

Syntactic dependency tree of a sentence

`http://epe.nlpl.eu/index.php?page=1`

A sequence of part-of-speech tags in a sentence:

DET	ADJ	NOUN	AUX	ADV	ADJ	PART	VERB	ADP	ADJ	NOUN	PUNCT
A	similar	technique	is	almost	impossible	to	apply	to	other	crops	.