

# Contradictory, My Dear Watson

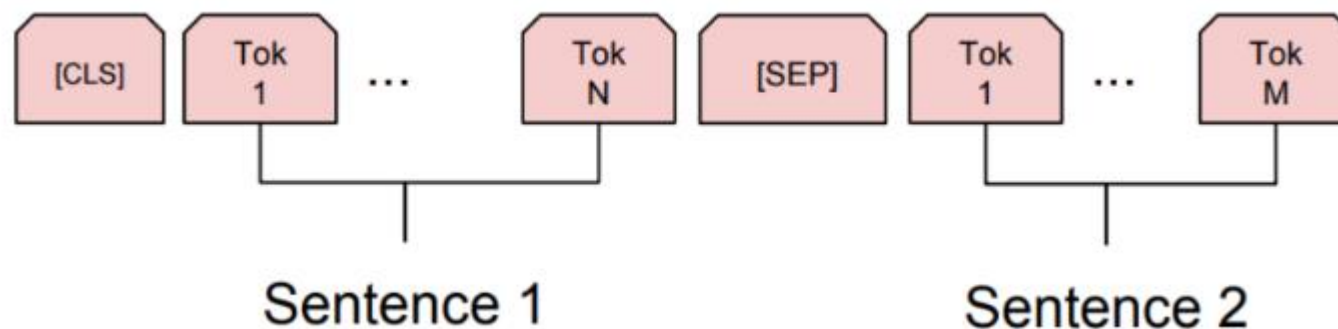
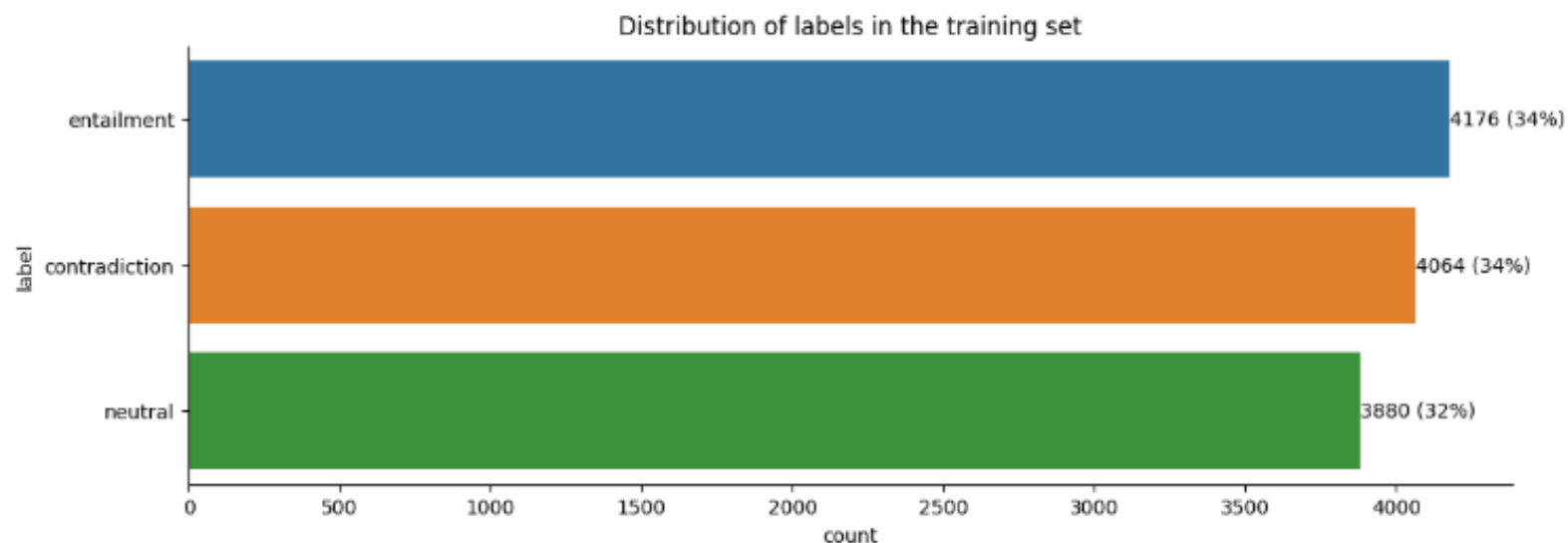
Midterm Report  
AI2 Seminar (NAIL052)

Martina Ciklamíniová

April 9, 2024

# Data

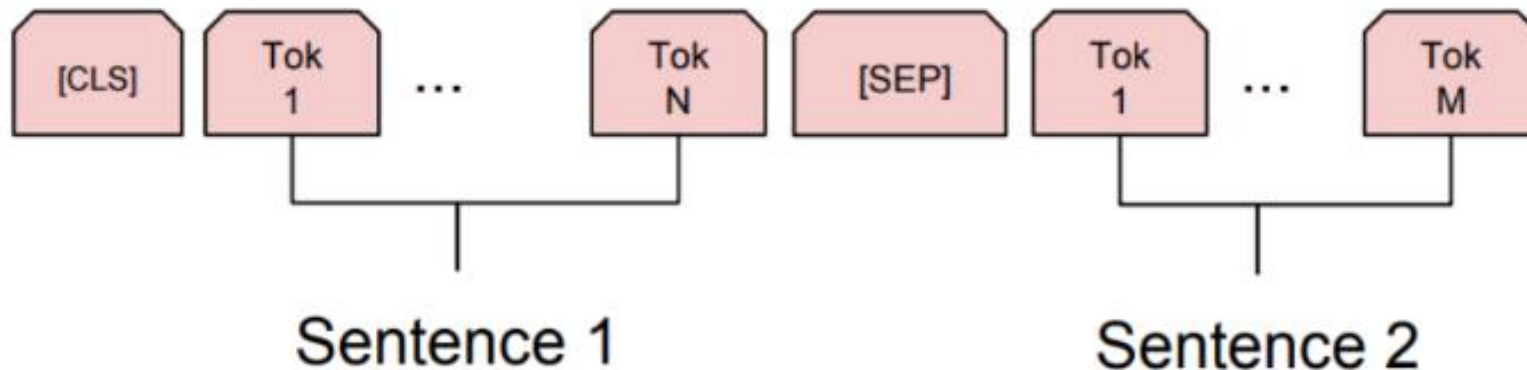
	hypothesis_length	premise_length
count	12120.000000	12120.000000
mean	15.316667	30.602145
std	7.637923	19.694049
min	1.000000	1.000000
25%	10.000000	16.000000
50%	14.000000	27.000000
75%	18.000000	41.000000
max	81.000000	232.000000



# Preprocessing data

- Preprocessor: "bert\_preprocessor"
- Special tokens "[CLS]", "[SEP]" and "[PAD]"

Tokenizer (type)	Vocab #
<code>bert_tokenizer (BertTokenizer)</code>	119,547



# Starting Point

Model: "bert\_classifier"

Layer (type)	Output Shape	Param #
padding_mask (InputLayer)	(None, None)	0
segment_ids (InputLayer)	(None, None)	0
token_ids (InputLayer)	(None, None)	0
bert_backbone (BertBackbone)	{sequence_output: (None, None, 768), pooled_output: (None, 768)}	177,853,440
dropout (Dropout)	(None, 768)	0
logits (Dense)	(None, 3)	2,307

Results:

- 0.61347
- 37th

- bert\_base\_multi
- 12-layer BERT model where case is maintained
- trained on Wikipedias of 104 languages

# Other multilingual models in KerasNLP

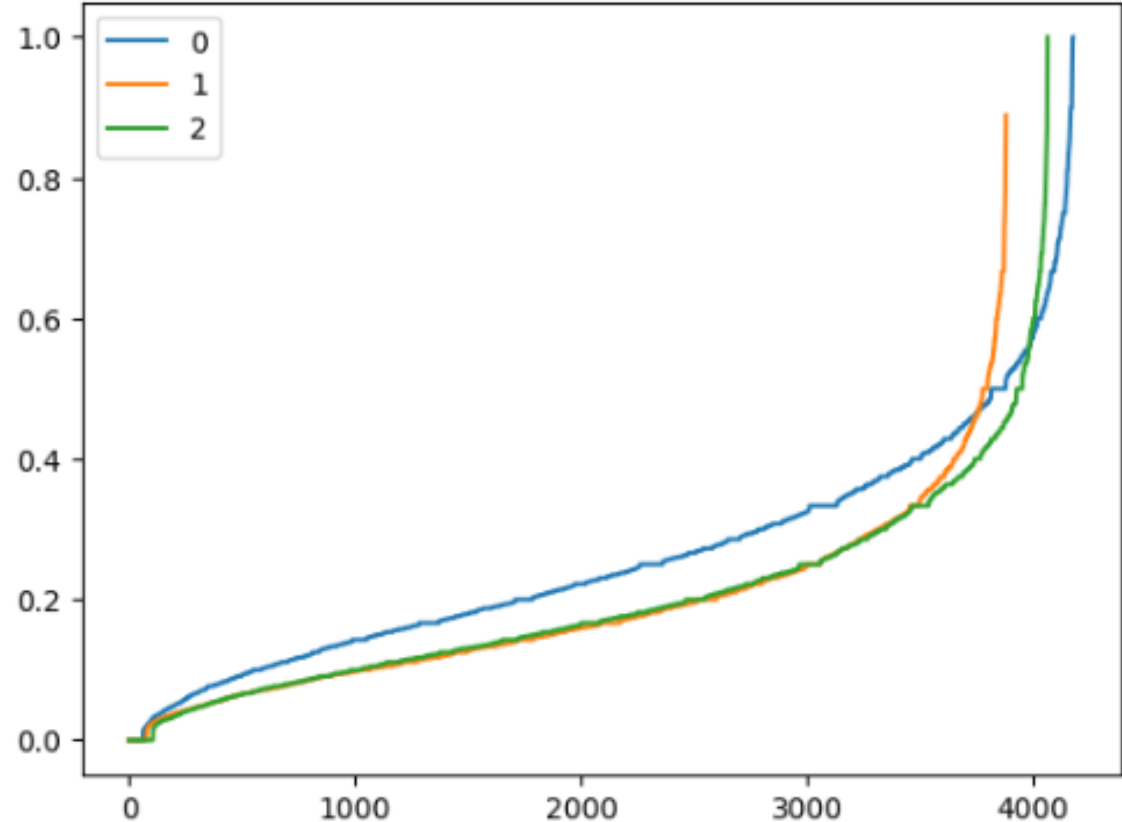
- bert\_base\_multi
- deberta\_v3\_base\_multi
- distil\_bert\_base\_multi
- xlm\_roberta\_base\_multi
- xlm\_roberta\_large\_multi

# Decision Tree

- 0.29412 – zero-padding at the end
- 0.33089 – zero-padded both sentences
- 0.35591 – with premise-hypothesis distance

# Premise-hypothesis distance

- Jaccardi similarity
  - the ratio of the cardinality of  $A \cap B$  and  $A \cup B$
- 0; entailment
  - mean: 0.2574
  - median: 0.2307
- 1; neutral
  - mean: 0.1807
  - median: 0.1559
- 2; contradiction
  - mean: 0.1940
  - median: 0.1666



# Future Work

- Figure out premise-hypothesis encoding
  - distinguishable features
- Look for more useful metrics
  - Decision tree only on metrics?