

Abstractive summarization of fact check reports with pre-trained transformer tuning on extractive summaries

Peter Vajdecka

vajp02@vse.cz

Content

Introduction to automated fact-checking

Extractive summarization

Abstractive summarization

Proposed process of hybrid summarization

Experiments

Introduction

- the problem of disinformation has been escalating in recent years (the controversies surrounding the US elections, covid-19, and now the war in Ukraine).

Disinformation crisis (global)



Fact-checking

Disinformation spread

- thanks to the international **Cimple** project for scientific and financial support on this work
([CIMPLE | CHIST-ERA \(chistera.eu\)](https://chistera.eu))

Cimple

Claim: 15 days left to vote BUT if you are voting by mail, you need to vote TODAY. USPS says it needs a 14 day roundtrip to be counted on election day.

Ruling comments: The United States is expected to break records for voting by mail this year and that's creating a deluge of claims on social media about deadlines. **Some are accurate, some are not.** "15 days left to vote BUT if you are voting by mail, you need to vote TODAY," reads one popular Instagram post. "USPS says it needs a 14 day roundtrip to be counted on election day." reads. This post was flagged as part of Facebook's efforts to combat false news and misinformation on its News Feed.(Read more about our partnership with Facebook.) The post is a screenshot of an Oct. 19 tweet by singer-songwriter Finneas Baird O'Connell, brother of singer Billie Eilish. Finneas' Twitter profile photo is of the Biden/Harris campaign logo, but he has no official role with the campaign. The post wrongly creates the impression that there is a national deadline to vote by mail. It's also a confusing message: the first sentence says voters must send it mail ballots 15 days ahead of time while the second sentence says the post office needs a 14-day roundtrip, suggesting that a voter can mail it in seven days ahead of time.

Here's what you should know:

The Postal Service did recommend in a national postcard in September that voters request the mail-in ballot at least 15 days before Election Day, Nov. 3, and return it at least seven days before Election Day. But this Instagram post omits the fact that states set their own laws about deadlines for receiving mail ballots. What's more, many states have options for voters to bypass the mail to return their ballots in an official ballot drop box or drop off site. Since the deadlines to return mail ballots vary by state, the best advice for voters is to check in with their local elections officials for information about when they must return their ballot, and their options for how to return it. Also, some states are automatically sending ballots to voters and therefore they don't have to request them. A spokesperson for the post office reiterated their previous advice, but also encouraged voters to check their state's requirements.

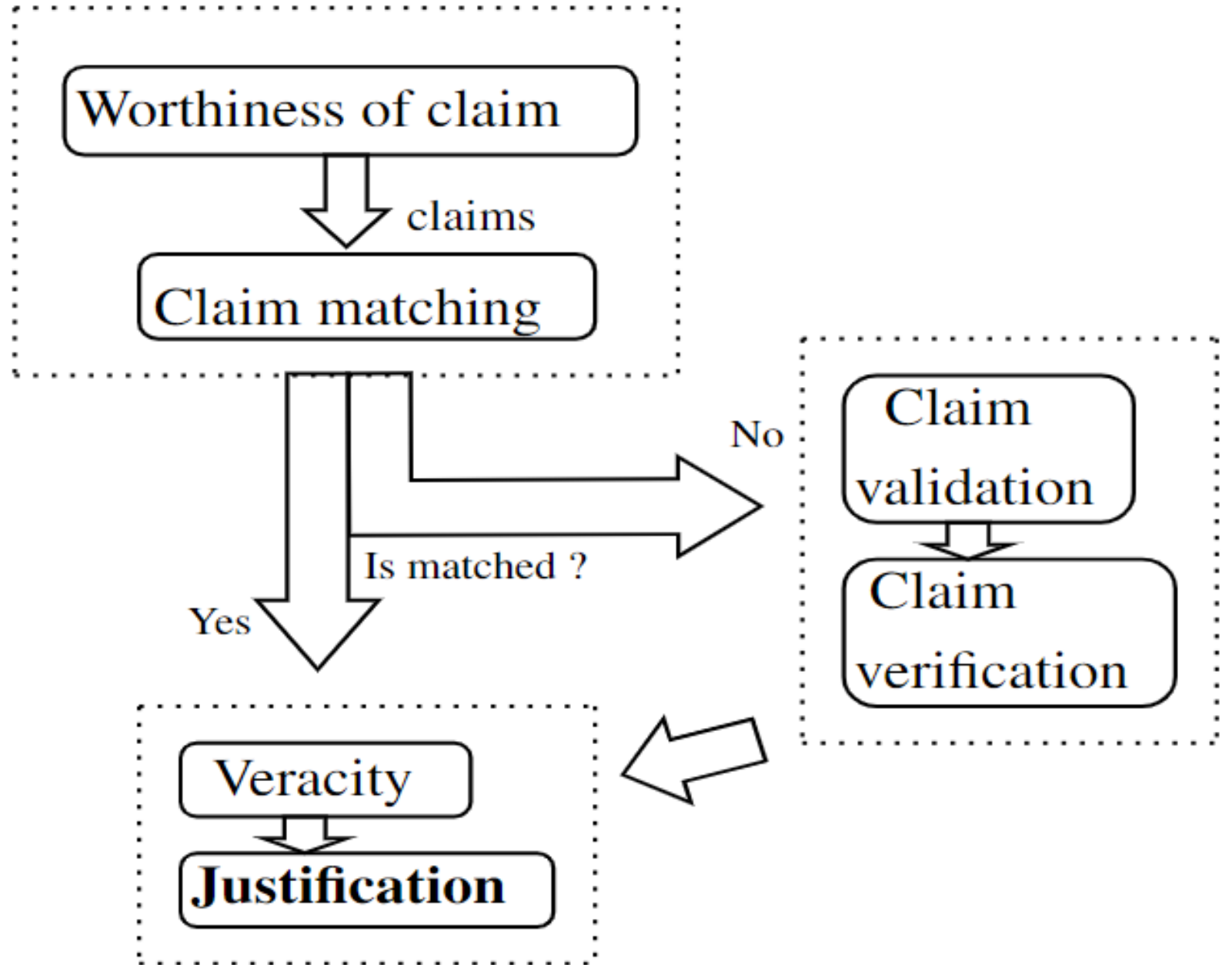
Justification: An Instagram post states "15 days left to vote BUT if you are voting by mail, you need to vote TODAY. USPS says it needs a 14 day roundtrip to be counted on election day. The post is unclear and omits important context. USPS said in a postcard in September that voters who want to have their ballots counted in the Nov. 3 general election should request the mail-in ballot at least 15 days before Election Day and return it at least seven days before Election Day. But states set deadlines for receiving mail ballots, and many jurisdictions allow voters to bypass the mail and return ballots in official ballot drop boxes or drop off sites. It's a good idea to return a ballot as soon as you can, but if you want to know the actual deadline for your state, check in with your state or local elections office. If you want to find out if your city or county has a place where you can drop it off, check in with your local elections office which typically posts that information on their website. ~~We rate this statement Half-True.~~

Veracity: Half-True

*EXAMPLE OF
A FACT-CHECKING REPORT*

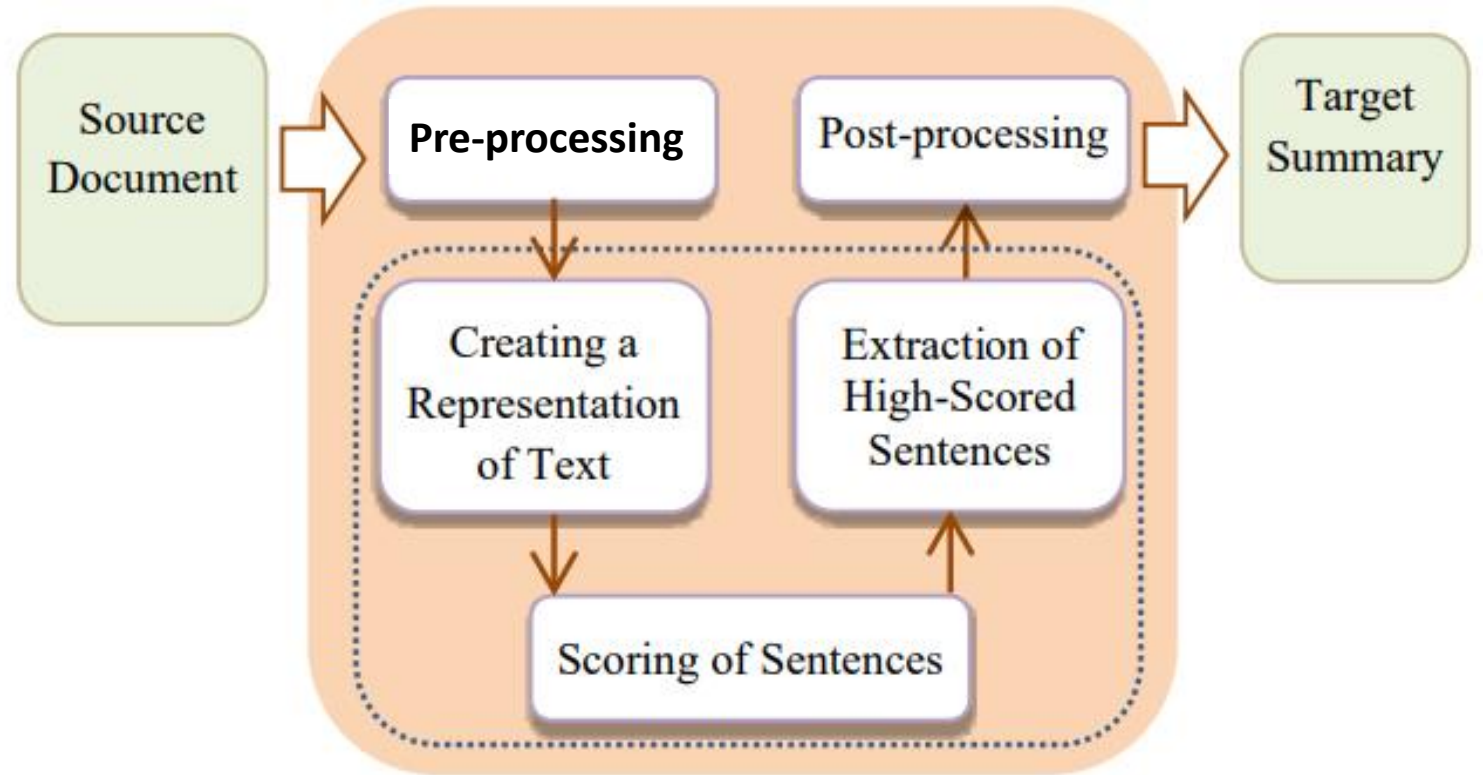
Automation of fact-checking

Claim Detection



EXTRACTIVE SUMMARIZATION

Extractive automatic summarization - Preprocessing



Text normalization

Sentence: Some are accurate, some are not.

Text normalization

Sentence: Some are accurate, some are not.

Tokens: ["Some", "are", "accurate", ",", "some", "are", "not", "."]

Text normalization

Sentence: Some are accurate, some are not.

Tokens: ["Some", "are", "accurate", ",", "some", "are", "not", "."]

NLP classification:

- Stop-words = „some“, „be“, „not“
- Punctuation = „“, „“
- Significant word = „accurate“

Text normalization

Sentence: Some are accurate, some are not.

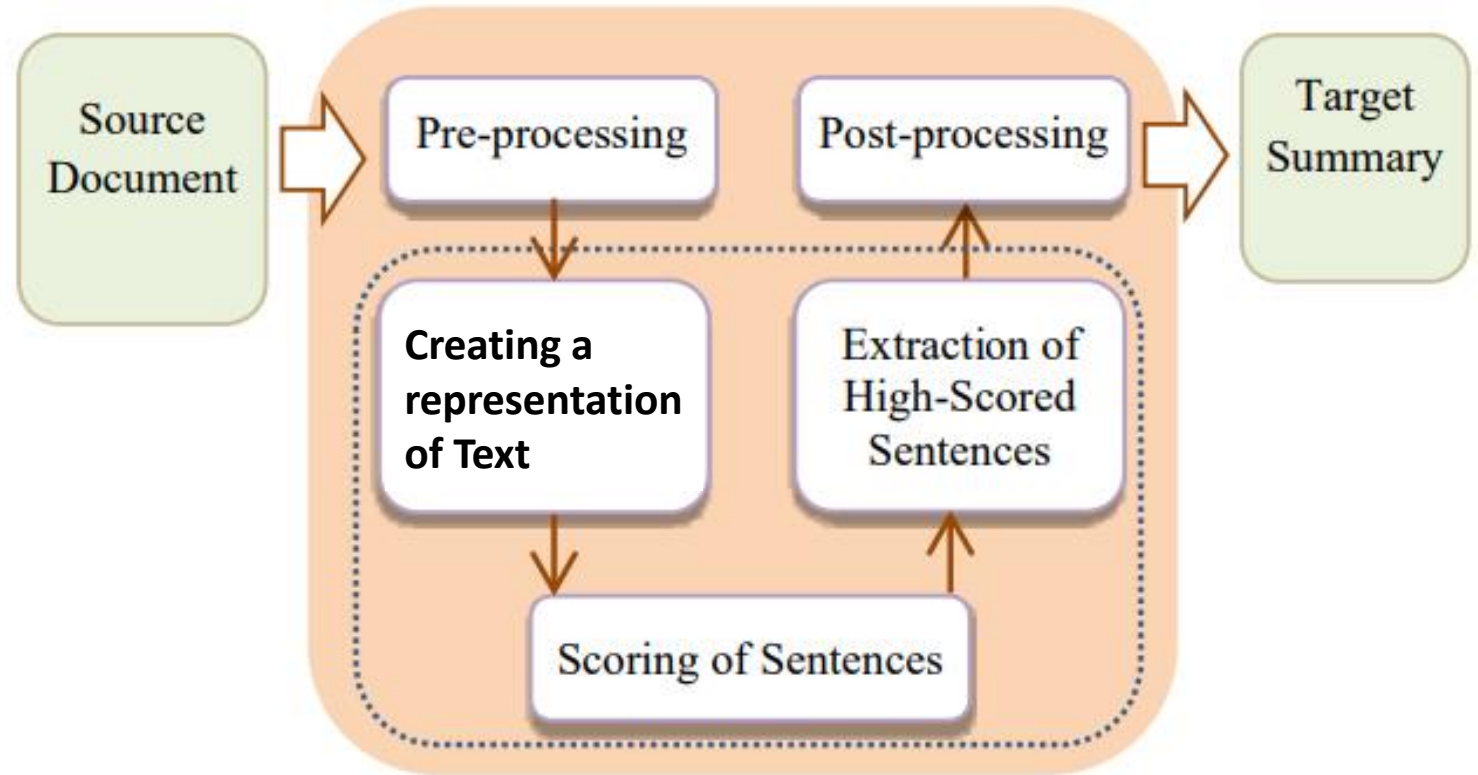
Tokens: ["Some", "are", "accurate", ",", "some", "are", "not", "."]

NLP classification:

- Stop-words = „some“, „be“, „not“
- Punctuation = „“, „“
- Significant word = „accurate“

Final tokens: ["accurate"]

Extractive automatic summarization - Vectorization



Ruling Comments	Order	Sentence	Sentence vector
The United States is expected to break records for voting mail this year and that's creating a deluge of claims on social media about deadlines. Some are accurate, some are not. 15 days left to vote BUT if you are voting mail, you need to vote TODAY, reads one popular Instagram post. USPS says it needs a 14 day roundtrip to be counted on election day. reads. This post was flagged as part of Facebook's efforts to combat false news and ...	1	The United States is expected to ...	$[w_{11}, w_{12}, \dots, w_{1n}]$
	2	Some are accurate, some are not.	$[w_{21}, w_{22}, \dots, w_{2n}]$
	3
	4
	5
	6
	7
	8	Finneas' Twitter profile photo is of ...	$[w_{81}, w_{82}, \dots, w_{8n}]$
	9
	10
	11
	12
	13
	14
	15	RELATED: Fact-checking the ...	$[w_{151}, w_{152}, \dots, w_{15n}]$
	16	A spokesperson for the post office ...	$[w_{161}, w_{162}, \dots, w_{16n}]$

Vectorization of sentences

1. Bag of words

s_1 : Some are accurate, some are not.

s_2 : Some are accurate.

s_3 : Some are not.

1. Bag of words

s_1 : Some are accurate, some are not.

s_2 : Some are accurate.

s_3 : Some are not.

	some	are	accurate	not
s_1	2	2	1	1
s_2	1	1	1	0
s_3	1	1	0	1

2. TF-IDF

$$\text{TF-IDF}(v, s) = \text{TF}(v, s) \times \text{IDF}(v, R)$$

$\text{TF}(v, s)$ = number of occurrences of word v in sentence s

$$\text{IDF}(v, R) = \log \left(\frac{|R|}{|\{s \in R : v \in s\}|} \right)$$

where : R = report (ruling comments)

v = word

s = sentence of report

2. TF-IDF

$$\text{TF-IDF}(v, s) = \text{TF}(v, s) \times \text{IDF}(v, R)$$

$\text{TF}(v, s)$ = number of occurrences of word v in sentence s

$$\text{IDF}(v, R) = \log \left(\frac{|R|}{|\{s \in R : v \in s\}|} \right)$$

where : R = report (ruling comments)

v = word

s = sentence of report

s_1 : Some are accurate, some are not.

s_2 : Some are accurate.

s_3 : Some are not.

2. TF-IDF

$$\text{TF-IDF}(v, s) = \text{TF}(v, s) \times \text{IDF}(v, R)$$

$\text{TF}(v, s)$ = number of occurrences of word v in sentence s

$$\text{IDF}(v, R) = \log \left(\frac{|R|}{|\{s \in R : v \in s\}|} \right)$$

where : R = report (ruling comments)
 v = word
 s = sentence of report

s_1 : Some are accurate, some are not.

s_2 : Some are accurate.

s_3 : Some are not.

	some	are	accurate	not
s_1	0	0	0.30	0.30
s_2	0	0	0.30	0
s_3	0	0	0	0.30

Table 3: TD-IDF weights

One-hot encoded vector

Vocabulary = {some, are, accurate, not}

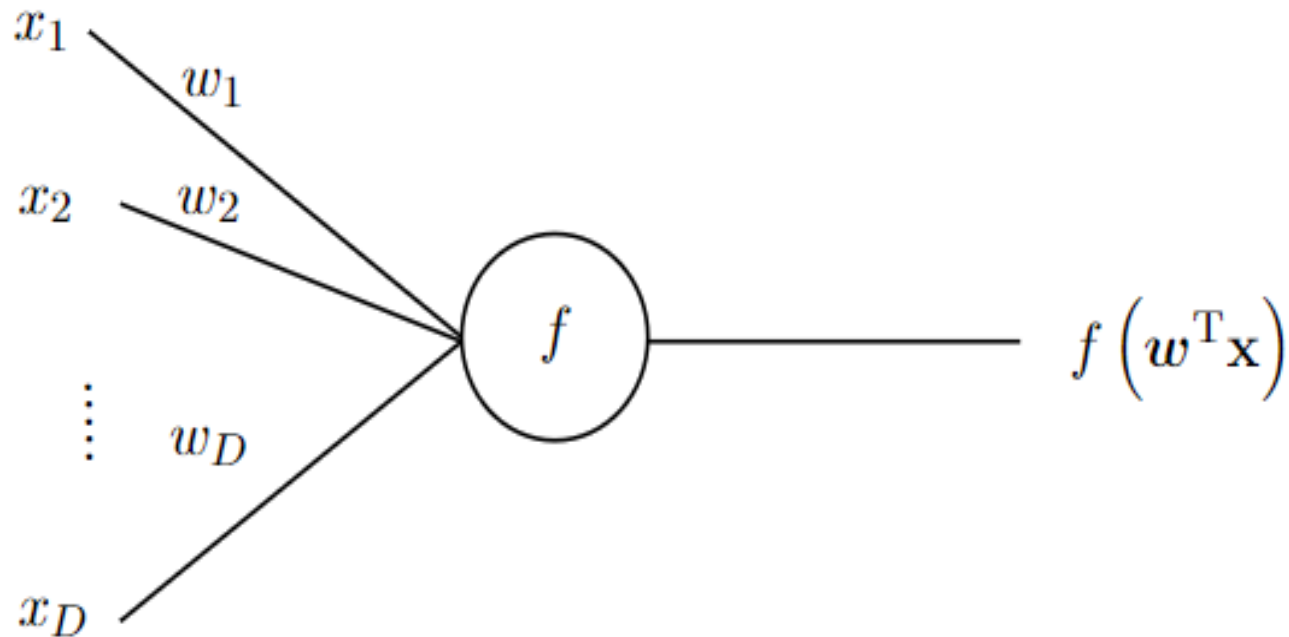
some = [1, 0, 0, 0]

are = [0, 1, 0, 0]

accurate = [0, 0, 1, 0]

not = [0, 0, 0, 1]

2. Word2vec - single neuron



- random weight (parameters or numbers)

1. $u = \mathbf{w}^T \mathbf{x}$.

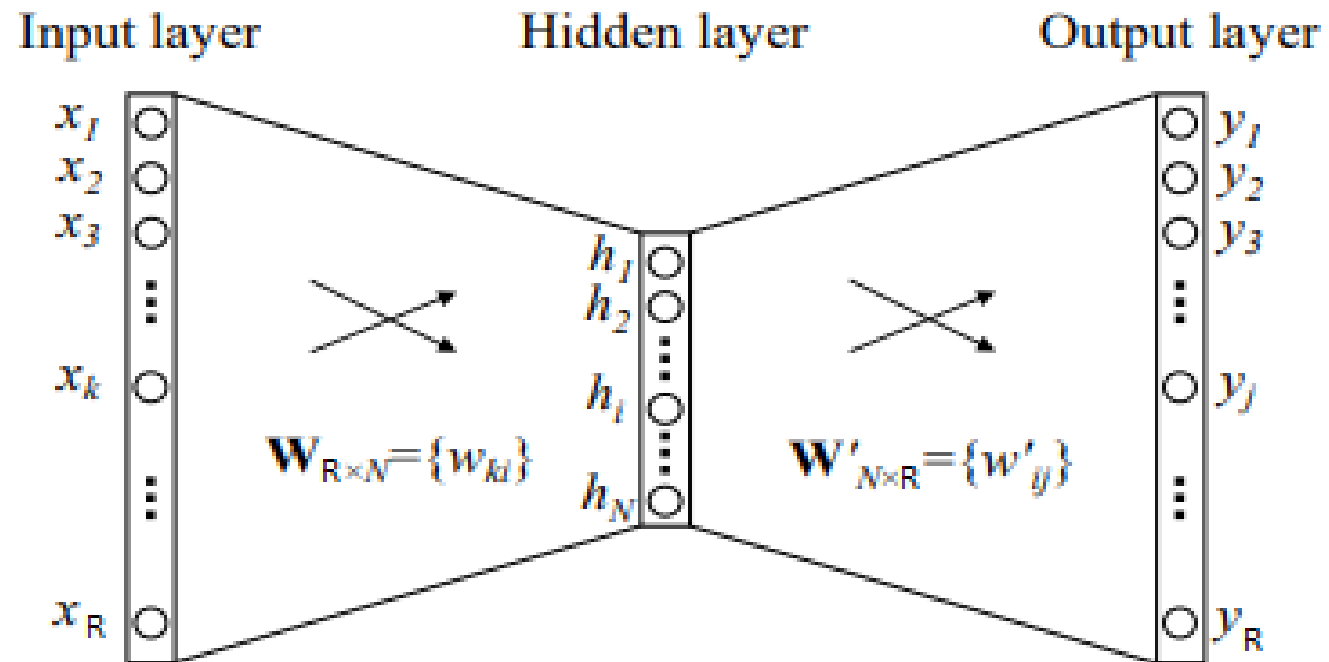
2. $y = \sigma(u) = \frac{1}{1 + e^{-u}}$

3. $E = \frac{1}{2}(t - y)^2$

Hint: Dot product of two vectors

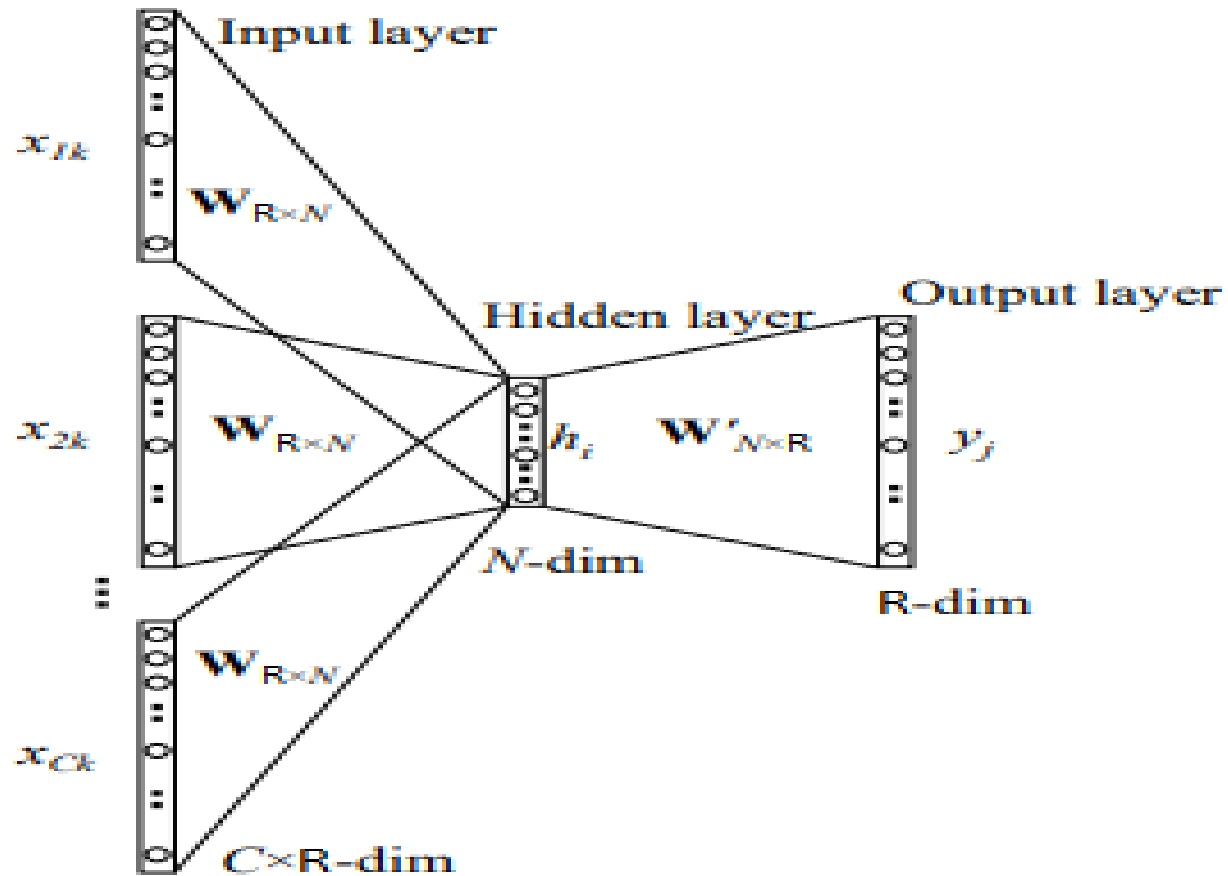
$$\begin{bmatrix} A_x & A_y & A_z \end{bmatrix} \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = A_x B_x + A_y B_y + A_z B_z = \vec{A} \cdot \vec{B}$$

2. Word2vec – three layers neural network



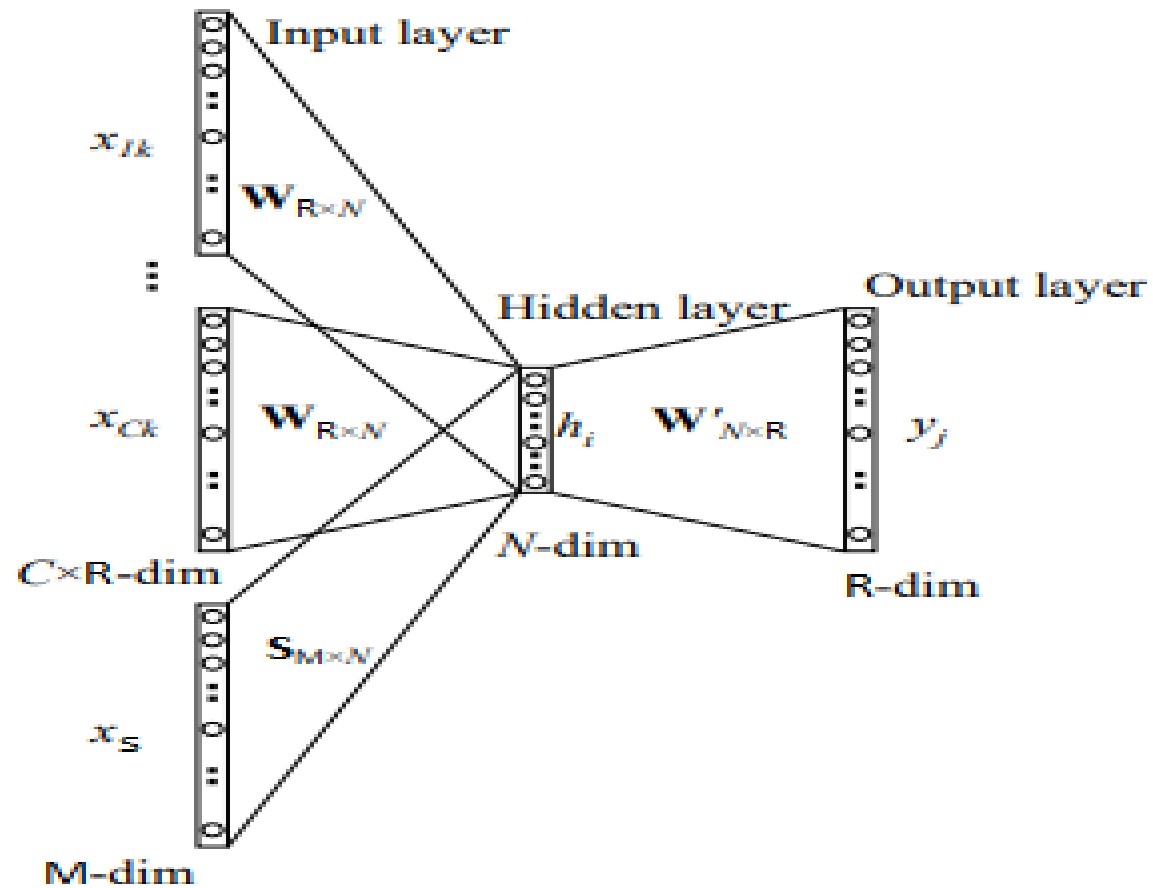
- one word as an input
- one word as an output
- two weight matrix

2. Word2vec – CBOW



- C words as an input
- one word as an output
- two weight matrix

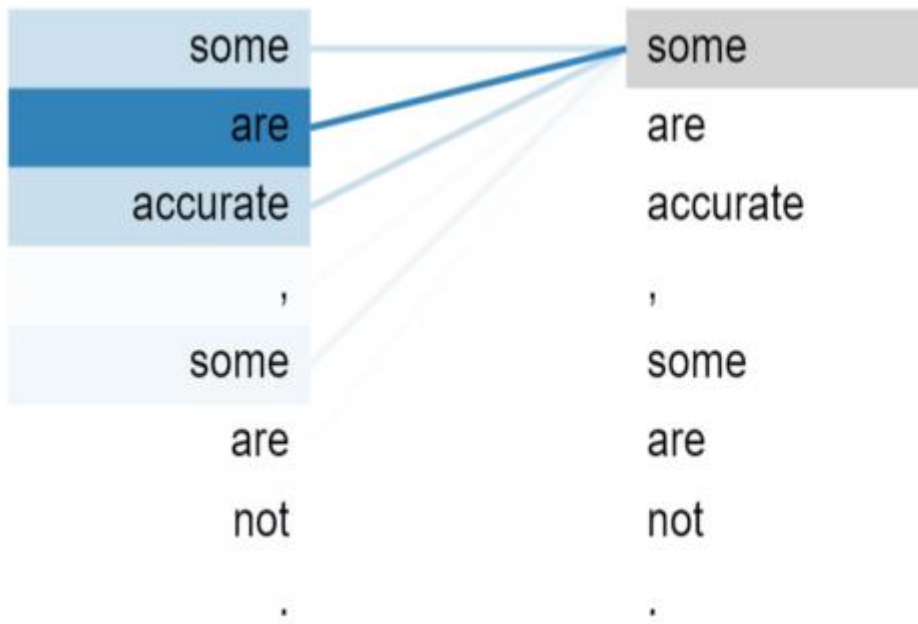
3. Doc2vec



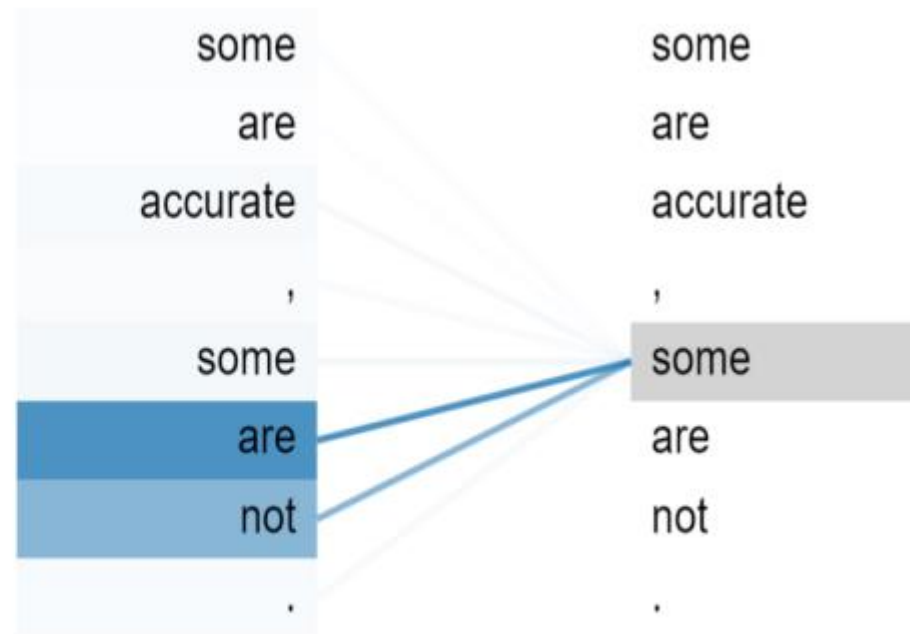
- C words as an input
- Sentence as an input
- one word as an output
- three weight matrix

4. BERT - Attention

- BERT= Bidirectional Encoder Representations from Transformers
- Attention = “How relevant is a token to others token and to itself”
- The darkness of the line determines the value of attention

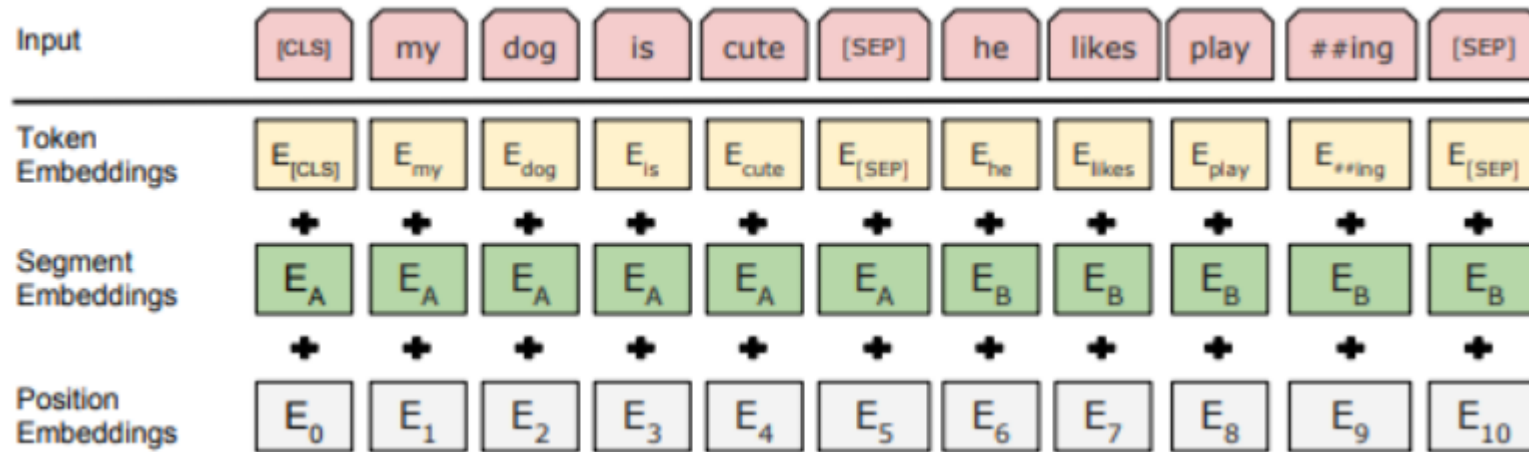


Attention to first token „some“



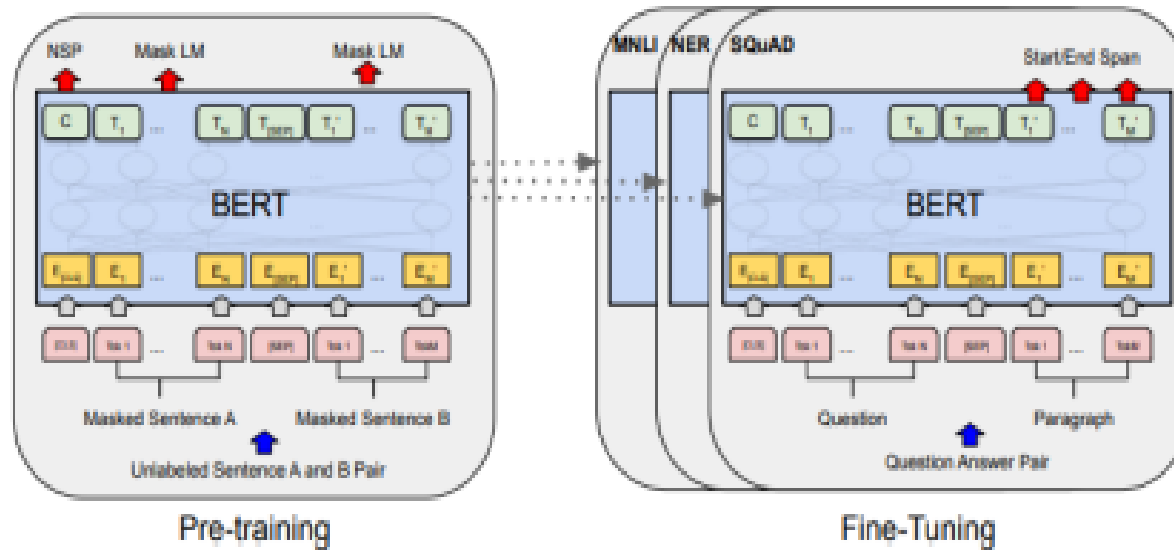
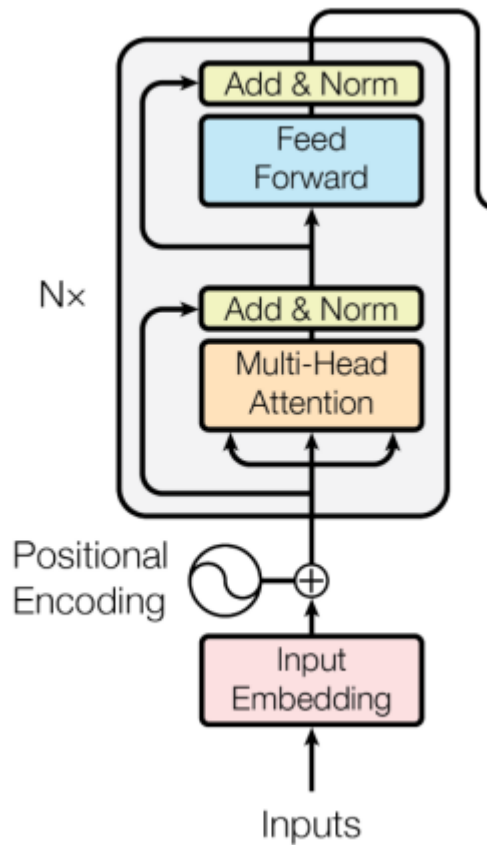
Attention to second token „some“

4. BERT - Representation



- vocabulary of 30 000 tokens (Wordpiece tokenization)
- two segment embeddings
- 512 token embeddings (after 512 token length, input is truncated)

4. BERT - architecture



- model with much more advanced architecture compared to doc2vec
- trained on prediction of next sentence and prediction of masked word
- fine-tuning entails retraining all parameters end-to-end

4. SENTENCE BERT

Much faster to fine-tune than BERT

Using siamese network (same neural network is shared)

Ideal for task of textual similarity

- **cosine similarity:**

$$d(\mathbf{u}, \mathbf{z}) = \cos(\mathbf{u}, \mathbf{z}) = \frac{\mathbf{u} \cdot \mathbf{z}}{\|\mathbf{u}\| \|\mathbf{z}\|}$$

Outperforms:

- universal sentence encoders
- baseline BERT:
 - vector of CLS token
 - mean pooling of all tokens in sentences
- elmo

4. Fine-tuning of SENTENCE BERT

We focus only on **Triplet Loss Objective Function** in our work for fine-tuning using:

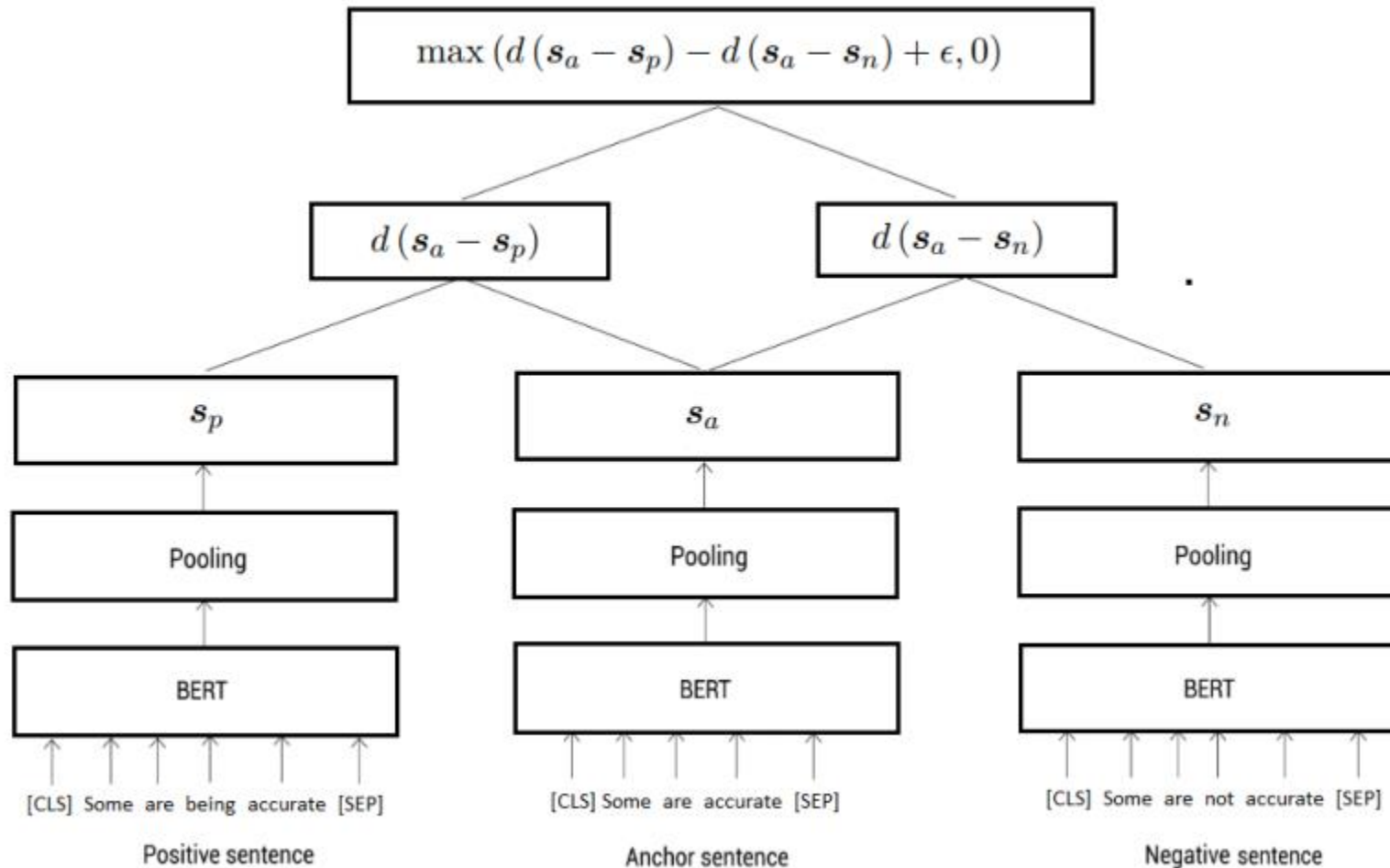
$$\max(d(\mathbf{s}_a - \mathbf{s}_p) - d(\mathbf{s}_a - \mathbf{s}_n) + \epsilon, 0)$$

Three types of inputs: anchor sentence, positive sentence, negative sentence

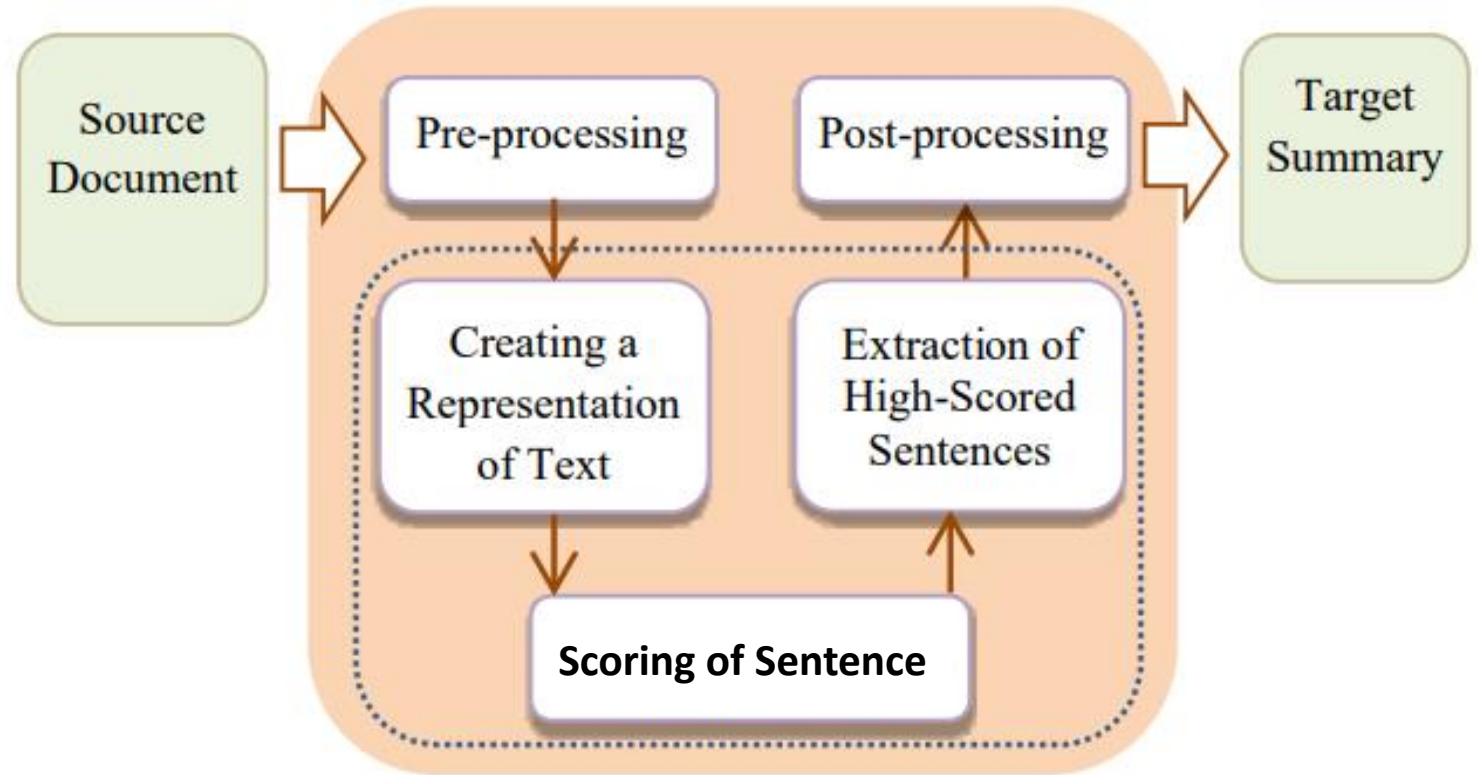
Each sentence vector comes from **same** Siamese network

We maximize distance between anchor and negative, but minimize distance between anchor and positive

4. Fine-tuning process of SENTENCE BERT



Extractive automatic summarization- Sentence Score

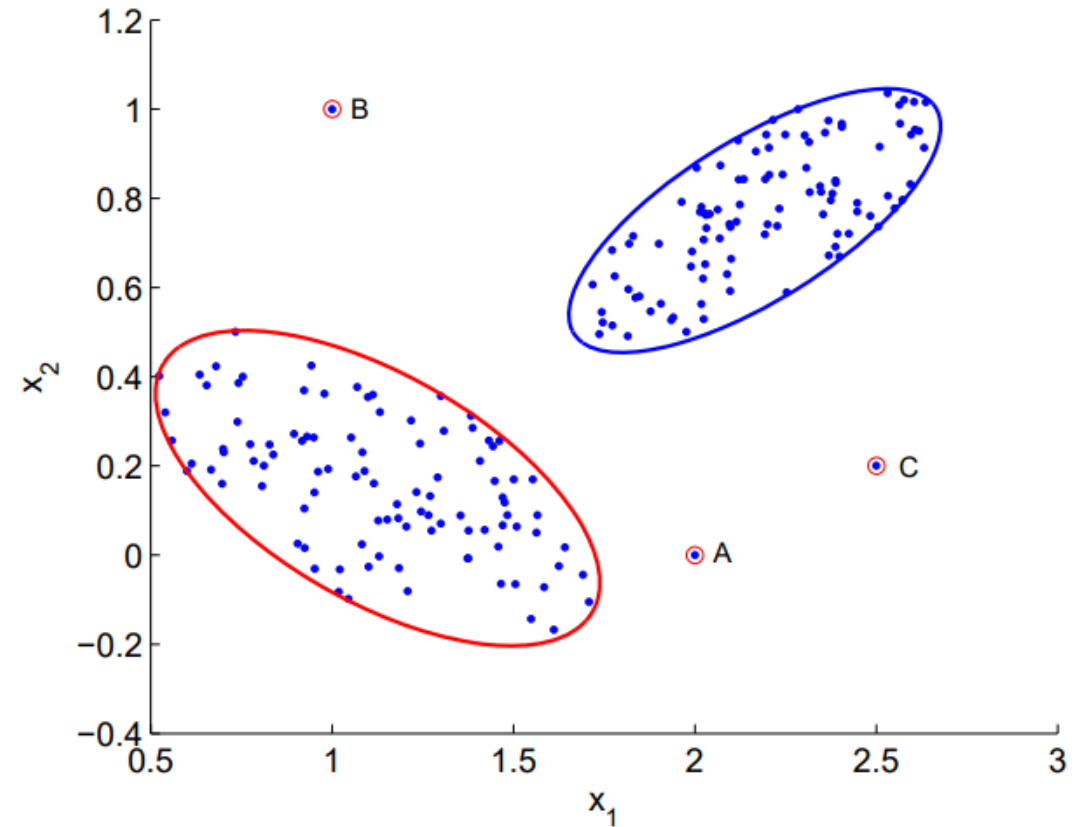


Local Outlier Factor (LOF)

Density based outlier detection

Each object (sentence) is assigned a local outlier factor

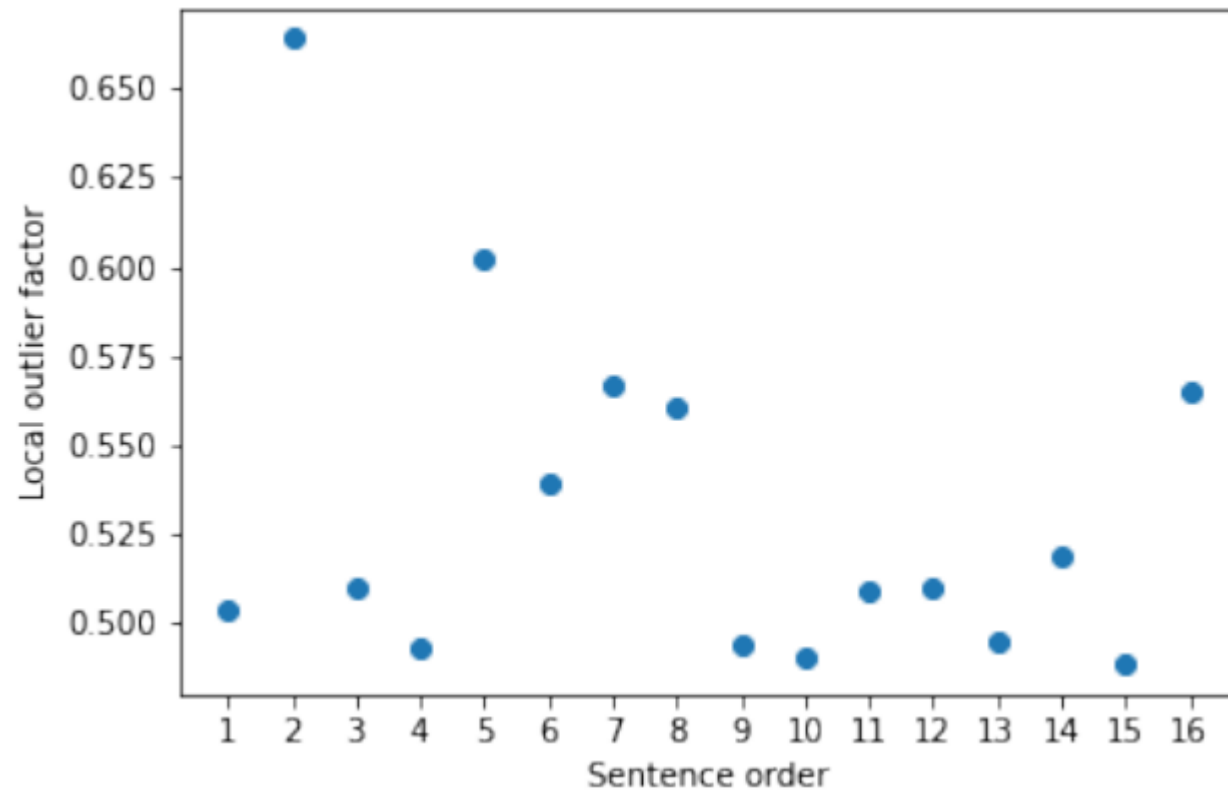
Object that have substantially lower density than their neighbors have high LOF



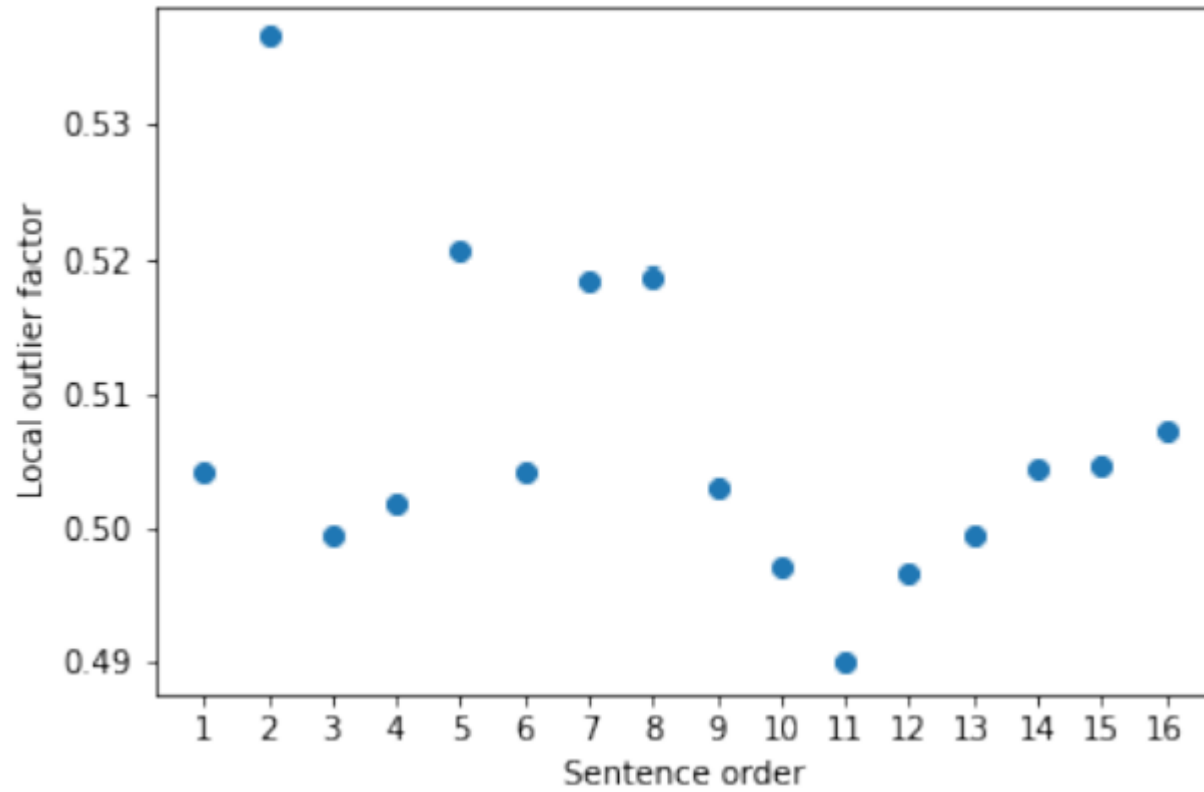
Ruling Comments	Order	Sentence	Sentence vector
The United States is expected to break records for voting mail this year and that's creating a deluge of claims on social media about deadlines. Some are accurate, some are not. 15 days left to vote BUT if you are voting mail, you need to vote TODAY, reads one popular Instagram post. USPS says it needs a 14 day roundtrip to be counted on election day. reads. This post was flagged as part of Facebook's efforts to combat false news and ...	1	The United States is expected to ...	$[w_{11}, w_{12}, \dots, w_{1n}]$
	2	Some are accurate, some are not.	$[w_{21}, w_{22}, \dots, w_{2n}]$
	3
	4
	5
	6
	7
	8	Finneas' Twitter profile photo is of ...	$[w_{81}, w_{82}, \dots, w_{8n}]$
	9
	10
	11
	12
	13
	14
	15	RELATED: Fact-checking the ...	$[w_{151}, w_{152}, \dots, w_{15n}]$
	16	A spokesperson for the post office ...	$[w_{161}, w_{162}, \dots, w_{16n}]$

Reminder of the report : LOF on sentence vectors

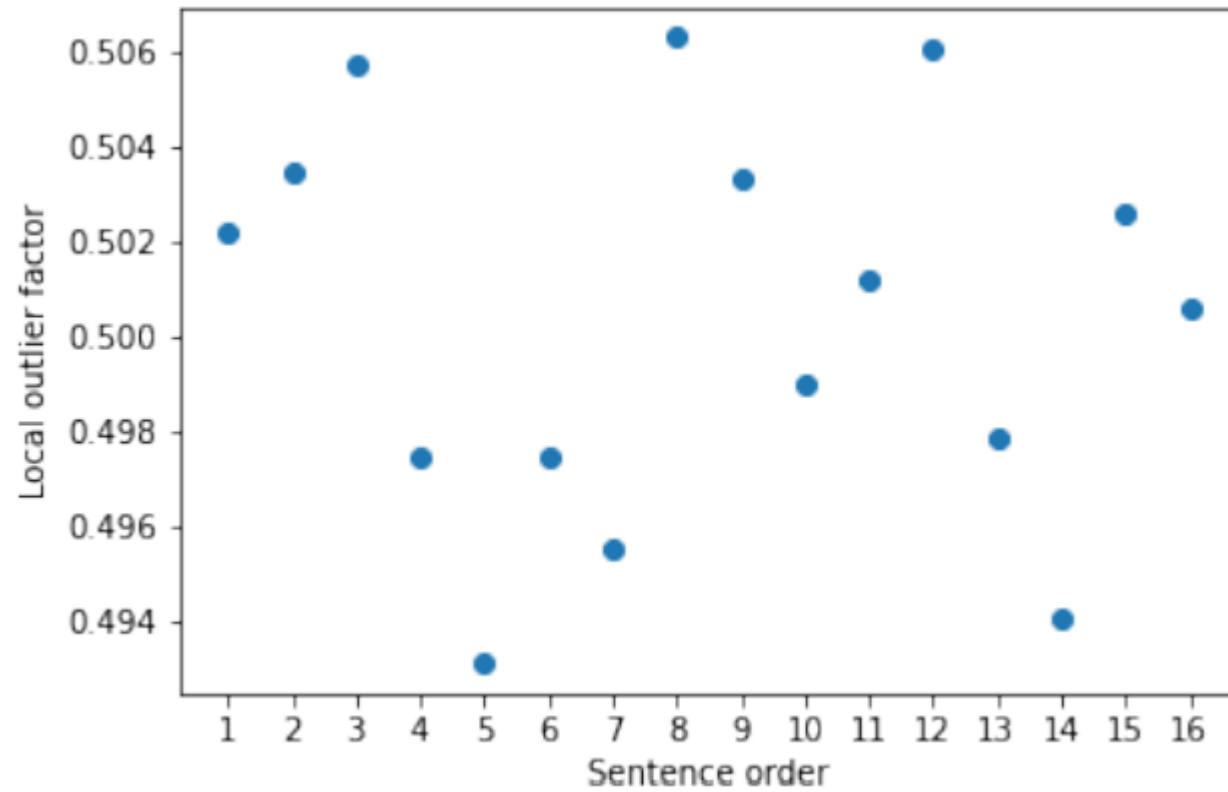
1. LOF + SENTENCE BERT



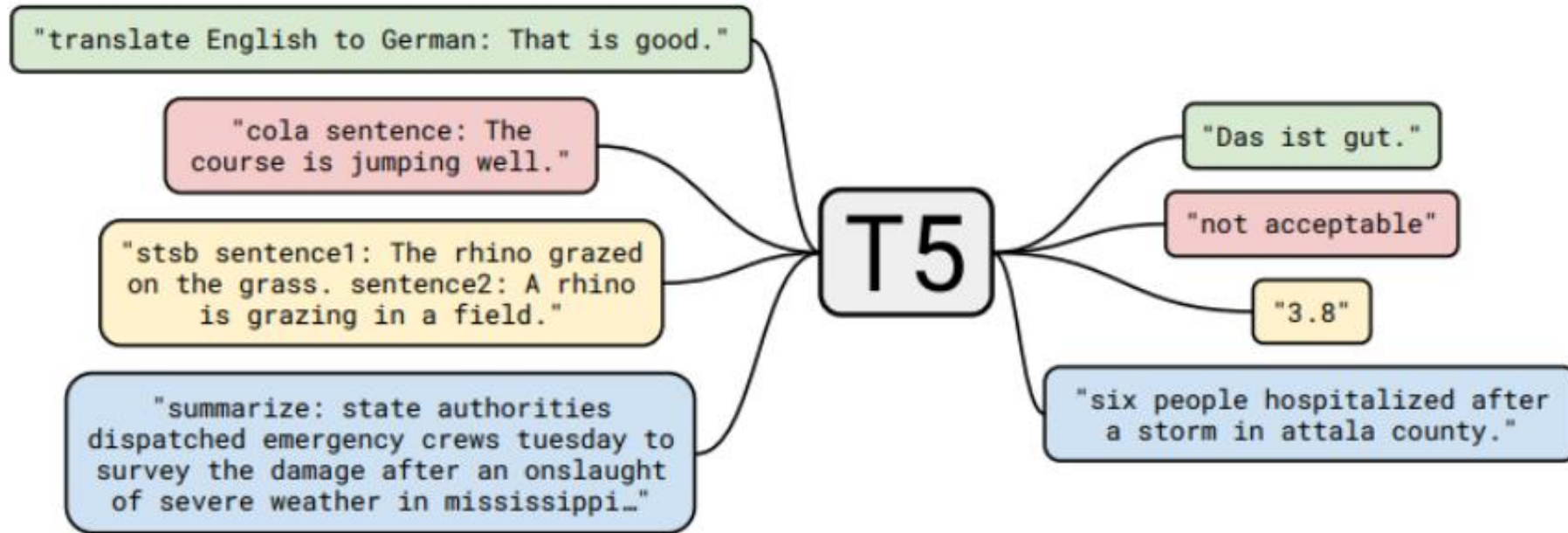
2. LOF + TF-IDF



3. LOF + Doc2Vec



ABSTRACTIVE SUMMARIZATION

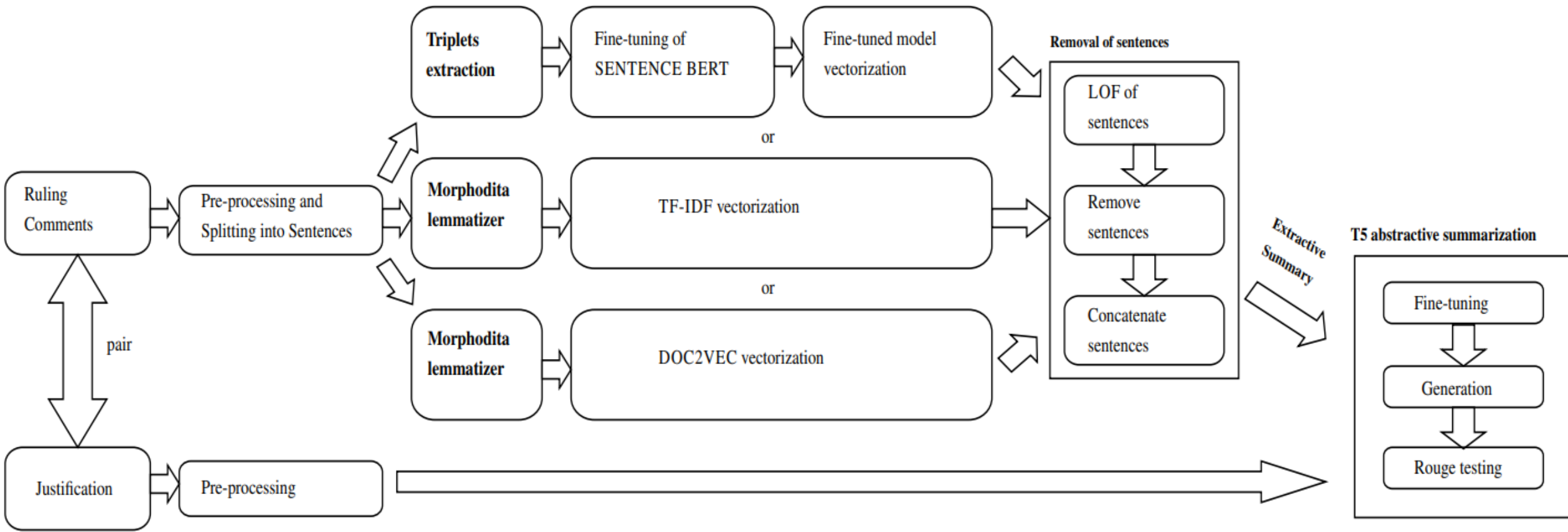


Unified text to text transformer (T5)

- trained on **clear** data
- created C4 dataset
- **encoder-decoder** system
- natural language generation model
- for generating a summary, it uses the prefix "summarize:"
- model fine-tuned on pairs (long text + summary)
- we use it as a **black-box**

THE PROPOSED PROCEDURE

Proposed hybrid summarization process



Triplets extraction

```
[1] Input:  $RC \leftarrow \{RC[1], \dots, RC[k]\}$ ,  $SM \leftarrow \{SM[1], \dots, RC[k]\}$ ,  $t \in \mathbb{Z}$ 
[2] Output:  $TRIPLETS \leftarrow \{TRIPLETS[1], \dots, TRIPLETS[t \cdot k]\}$ 
[3] for  $i \leftarrow 1$  to  $k$  do
[4]    $SMVEC \leftarrow \emptyset$ 
[5]    $RCSPLIT \leftarrow \emptyset$ 
[6]    $SVEC \leftarrow \emptyset$ 
[7]    $POS \leftarrow \emptyset$ 
[8]    $NEGATIVE \leftarrow \emptyset$ 
[9]    $COS \leftarrow \emptyset$ 
[10]   $n \leftarrow 0$ 
[11]   $SMVEC \leftarrow \mathbf{sbert}(SM[i])$  // vectorize i-th summary
[12]   $RCSPLIT \leftarrow \mathbf{split}(RC[i])$  // split i-th report into sentences
[13]   $n \leftarrow |RCSPLIT|$ 
[14]  for  $j \leftarrow 1$  to  $n$  do
[15]     $SVEC[j] \leftarrow \mathbf{sbert}(RCSPLIT[j])$  // vectorize j-th sentence in i-th
      report
[16]     $COS[j] \leftarrow \mathbf{d}(SVEC[j], SMVEC)$  // calculate cosine distance
      between j-th sentence in i-th report and i-th summary
[17]  end
[18]   $RCSPLIT_{desc} \leftarrow \mathbf{orderdesc}(RCSPLIT, COS)$  // order sentences in i-th
      report by cosine similarity
[19]   $POS \leftarrow \mathbf{top}(RCSPLIT_{desc}, t)$  // t most similar sentences to summary
[20]   $NEGATIVE \leftarrow \mathbf{tail}(RCSPLIT_{desc}, t)$  // t least similar sentences to
      summary
[21]  foreach  $anchor \in POS$  do
[22]     $POSITIVE \leftarrow \emptyset$ 
[23]     $positive \leftarrow \emptyset$ 
[24]     $negative \leftarrow \emptyset$ 
[25]     $POSITIVE \leftarrow POS \setminus anchor$  // remove anchor from positives
[26]     $positive \leftarrow \mathbf{random}(POSITIVE)$ 
[27]     $negative \leftarrow \mathbf{random}(NEGATIVE)$ 
[28]     $TRIPLETS \leftarrow TRIPLETS \cup (anchor, positive, negative)$  // save each
      new triplet
[29]  end
[30] end
```


Extractive summarization using LOF

```
[1] Input:  $RC \leftarrow \{RC[1], \dots, RC[k]\}$ ,  $p \in (0, 1)$ 
[2] Output:  $ERC \leftarrow \{ERC[1], \dots, ERC[k]\}$ 
[3]  $RCSPLIT \leftarrow \emptyset$ 
[4]  $RC_{concatenated} \leftarrow \emptyset$ 
[5]  $RCSPLIT_{LOF_{desc}} \leftarrow \emptyset$ 
[6]  $RCSPLIT_{SORDER_{desc}} \leftarrow \emptyset$ 
[7]  $RCSPLIT_{extracted} \leftarrow \emptyset$ 
[8]  $SORDER \leftarrow \emptyset$ 
[9]  $SVEC \leftarrow \emptyset$ 
[10]  $SVEC_{desc} \leftarrow \emptyset$ 
[11]  $LOF \leftarrow \emptyset$ 
[12] for  $i \leftarrow 1$  to  $k$  do
[13] |  $n \leftarrow 0$ 
[14] |  $RCSPLIT[i] \leftarrow \text{split}(RC[i])$  // split i-th report into sentences
[15] |  $n \leftarrow |RCSPLIT[i]|$  // get number of sentences
[16] | for  $j \leftarrow 1$  to  $n$  do
[17] | |  $SORDER[i][j] \leftarrow j$ 
[18] | |  $SVEC[i][j] \leftarrow \text{vectorize}(RCSPLIT[i][j])$  // vectorize j-th sentence
[19] | end
[20] | for  $j \leftarrow 1$  to  $n$  do
[21] | |  $LOF[i][j] \leftarrow LOF_{Norm}(SVEC[i], SVEC[i][j])$  // calculate  $LOF_{Norm}$  for
[22] | | j-th sentence
[23] | end
[24] |  $RCSPLIT_{LOF_{desc}} \leftarrow \text{orderalldesc}(RCSPLIT, LOF)$  // order all sentences in
[25] | dataset by their  $LOF_{Norm}$ 
[26] |  $RCSPLIT_{extracted} \leftarrow \text{deletepercentage}(RCSPLIT_{LOF_{desc}}, p)$  // delete p
[27] | percentage of sentences from dataset
[28] |  $RCSPLIT[i]_{SORDER_{desc}} \leftarrow \text{orderasc}(RCSPLIT[i]_{extracted}, SORDER[i])$  // sort
[29] | the sentences according to their ordering in the initial
[30] | report
[31] |  $RC[i]_{concatenated} \leftarrow \text{concat}(RCSPLIT[i]_{SORDER_{desc}})$  // create plain text
[32] | from sentences
[33] | end
[34] if  $|RC_{concatenated}| \neq k$  then
[35] | break
[36] else
[37] |  $ERC \leftarrow RC_{concatenated}$  // save extractive summary of i-th report
[38] end
```

EXPERIMENTS

Evaluation metrics

T5 transformer is evaluated using **ROUGE** (english) and **ROUGERAW**(Czech)

Two variants **ROUGE-N** or **ROUGE-L**

ROUGE-N (or **ROUGERAW-N**) quantifies amount of overlap between generated and reference summaries in terms of **N-grams**

ROUGE-L (or **ROUGERAW-L**) examines the reference and the generated summaries for the longest subsequencies

ROUGE-1 (or **ROUGERAW-1**) , **ROUGE-1** (or **ROUGERAW-2**) , **ROUGE-L** (or **ROUGERAW-L**) were chosen to evaluate the approaches we suggest

Experimental framework

Used GPU server graphics cards **Nvidia A100** and **Nvidia A40** for fine-tuning SENTENCE BERT and T5 transformer

For English, we used **T5 Base** with **220M** parameters

For Czech, we apply **mT5 Base** with **580M** parameters (trained on 101 languages)

The code (without data and models) is publicly available in my github repository:

[petervajdecka02947/MasterThesis2022](https://github.com/petervajdecka02947/MasterThesis2022)

Politifact dataset

Split 80/10/10 = 10312/1289/1290

Dataset scraping from politifact.com:

- only pages containing "Our ruling" or "Our Rating" were considered, ensuring that the justification was manually created
- then characters like „\n“ or „\t“ were removed for text fluency,
- we have removed the html tags,
- urls have been removed,
- remove all sentences from the justification that contain words related to truthfulness of the claim,
- we have replaced 2 or more spaces with a single space,
- we updated all data from past until 17-th of February 2022.

	Ruling comments length	Justification length
count	12891.00	12891.00
mean	793.71	85.51
std	289.86	42.38
min	40.00	2.00
25%	589.00	58.00
50%	755.00	80.00
75%	953.00	106.00
max	2935.00	1121.00

Czech datasets

Demagog dataset

Split 80/10/10 = 2724/431/431

	Ruling comments length	Justification length
count	3406.00	3406.00
mean	295.45	30.26
std	179.22	9.92
min	32.00	4.00
25%	169.00	23.00
50%	253.00	31.00
75%	380.00	38.00
max	1531.00	95.00

SumeCzech dataset

Training = 77866

Validation = 44567

Testing = 44454

Out of Domain test = 44967

	Text length	Headline length
count	211863.00	211863.00
mean	401.38	8.76
std	307.25	2.45
min	99.00	3.00
25%	224.00	7.00
50%	319.00	9.00
75%	473.00	11.00
max	13283.00	22.00

Results - Politifact

Politifact.com					
Source	System	Rouge 1	Rouge 2	Rouge L	
Atanasova 2020 (University of Copenhagen)	Explain-Extractive	35.7	13.51	31.58	
	Explain-MT	35.13	12.9	30.93	
Kazemi 2021 (University of Michigan)	TextRank	27.74	7.42	23.24	
	GPT-2	24.01	5.78	21.15	
	Biased TextRank	30.90	10.39	26.22	
Vajdecka 2022 (VSE)	T5 Baseline	38.12	18.90	35.71	
	SBERT+ LOF+T5 (13 % of sentences removed)	38.35	18.88	35.88	
	Claim + T5 Baseline	39.19	20.56	36.92	
	CLAIM + SBERT+LOF+T5 (13 % of sentences removed)	39.45	21.08	37.27	
	CLAIM + SBERT+LOF+T5 (11 % of sentences removed)	39.76	21.37	37.54	
	CLAIM + SBERT fine-tuned +LOF+T5 (13 % of sentences removed)	40.76	22.00	38.36	
	CLAIM + SBERT fine-tuned +LOF+T5 (11 % of sentences removed)	39.55	20.69	37.11	
	CLAIM + MORPHODITA + TF-IDF+LOF+T5 (13 % of sentences removed)	39.91	20.62	37.40	
	CLAIM +MORPHODITA + TF-IDF+LOF+T5 (11 % of sentences removed)	39.86	20.59	37.30	
	CLAIM + MORPHODITA + DOC2VEC+LOF+T5 (13 % of sentences removed)	38.58	19.62	36.20	
	CLAIM + MORPHODITA + DOC2VEC+LOF+T5 (11 % of sentences removed)	39.04	20.65	36.70	

Results - Demagog

System	Test set								
	ROUGE _{RAW-1}			ROUGE _{RAW-2}			ROUGE _{RAW-L}		
	P	R	F	P	R	F	P	R	F
T5 Baseline	31.10	17.84	21.53	11.38	6.54	7.83	24.78	14.42	17.29
Claim + T5 Baseline	31.16	18.35	22.08	11.80	6.79	8.23	24.80	14.86	17.73
Claim + SBERT + LOF + T5 (24 % of sentences removed)	31.95	17.33	21.43	12.01	6.31	7.82	25.30	13.85	17.04
Claim + SBERT fine-tuned + LOF + T5 (24 % of sentences removed)	32.73	18.75	22.66	12.97	7.23	8.82	26.29	15.11	18.25
Claim + TF-IDF + LOF + T5 (24 % of sentences removed)	30.58	19.92	23.08	11.70	7.51	8.74	24.03	15.82	18.24
Claim + DOC2VEC + LOF + T5 (24 % of sentences removed)	31.41	16.89	20.82	11.50	6.06	7.49	25.29	13.78	16.89

Results – SumeCzech

Text → Headline

Source	System	Test set									Out-of-domain test set								
		$\text{ROUGE}_{\text{RAW}-1}$			$\text{ROUGE}_{\text{RAW}-2}$			$\text{ROUGE}_{\text{RAW}-L}$			$\text{ROUGE}_{\text{RAW}-1}$			$\text{ROUGE}_{\text{RAW}-2}$			$\text{ROUGE}_{\text{RAW}-L}$		
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
SumeCzech (Straka et al., 2018)	first	7.4	13.5	8.9	1.1	2.2	1.3	6.5	11.7	7.7	6.7	13.6	8.3	1.3	2.8	1.6	5.9	12.0	7.4
	random	5.9	10.3	6.9	0.5	1.0	0.6	5.2	8.9	6.0	5.2	10.0	6.3	0.6	1.4	0.8	4.6	8.9	5.6
	textrank	6.0	16.5	8.3	0.8	2.3	1.1	5.0	13.8	6.9	5.8	16.9	8.1	1.1	3.4	1.5	5.0	14.5	6.9
	tensor2tensor	8.8	7.0	7.5	0.8	0.6	0.7	8.1	6.5	7.0	6.3	5.1	5.5	0.5	0.4	0.4	5.9	4.8	5.1
Named entities (Marek et al., 2021)	Seq2Seq	16.1	14.1	14.6	2.5	2.1	2.2	14.6	12.8	13.2	13.1	11.8	12	2	1.7	1.8	12.1	11	11.2
	Seq2Seq-NER	16.2	14.1	14.7	2.5	2.1	2.2	14.7	12.8	13.3	13.7	11.9	12.4	2	1.7	1.8	12.6	11.1	11.4
Peter Vajdecka (only 10 % of training data)	T5	15.4	11.0	12.5	3.2	2.3	2.6	14.2	10.1	11.5	15.9	11.9	13.2	4.4	3.2	3.6	14.9	11.2	12.4
	T5-SBERT-LOF	15.8	11.4	12.9	3.5	2.5	2.8	14.6	10.6	11.9	16.5	12.4	13.7	4.8	3.5	3.9	15.4	11.6	12.9

- 16 % of sentences removed from whole dataset with the best model

Politifact – the best summaries

Claim	Justification	Automatic summary
<p>Look at the men on this stage. Collectively, they have lost 10 elections. The only people on this stage who have won every single election that they've been in are the women, Amy and me.</p>	<p>Warren said, Look at the men on this stage. Collectively, they have lost 10 elections. The only people on this stage who have won every single election that they've been in are the women, Amy and me. We checked the electoral histories of the candidates on stage and found that Warren spoke accurately.</p>	<p>Warren said, Collectively, they have lost 10 elections. The only people on this stage who have won every single election that they've been in are the women, Amy and me.</p>
<p>Says there have been some job gains in the McMansion State since Mr. Christie took office, but they have lagged gains both in the nation as a whole and in New York and Connecticut, the obvious points of comparison.</p>	<p>In his May 28 column, Krugman claimed "there have been some job gains in the McMansion State since Mr. Christie took office, but they have lagged gains both in the nation as a whole and in New York and Connecticut, the obvious points of comparison." Krugman's analysis is correct. In terms of the percentage increases in total jobs and private-sector jobs as of April 2012, New Jersey's growth since the beginning of Christie's tenure has been less than the increases in New York, Connecticut and the nation as a whole. To comment on this ruling, go to NJ.com.</p>	<p>Krugman said, "There have been some job gains in the McMansion State since Mr. Christie took office, but they have lagged gains both in the nation as a whole and in New York and Connecticut, the obvious points of comparison." In terms of total jobs and private-sector job growth since January 2010, New Jersey's overall job growth has been less than New York, Connecticut and the nation as a whole during the same period. To comment on this ruling, go to NJ.com.</p>
<p>Two Iraqis came here to this country, were radicalized and they were the masterminds behind the Bowling Green massacre. ... It didn't get covered</p>	<p>Conway said that "two Iraqis came here to this country, were radicalized and they were the masterminds behind the Bowling Green massacre. ... It didn't get covered." Put simply, there was no massacre.</p>	<p>Conway said that "two Iraqis came here to this country, were radicalized and they were the masterminds behind the Bowling Green massacre.... It didn't get covered." There was indeed a terrorism-related situation in Bowling Green, Ky., but it wasn't a massacre.</p>

Politifact – the worst summaries

Claim	Justification	Automatic summary
In South Dakota, I'm putting prayer back in our schools.	No specific policy exists, nor does any specific proposal or a timeline for producing one. That could change if Noem produces a proposal, the Legislature approves it, and it passes muster in court.	Noem told the crowd, In South Dakota, I'm putting prayer back in our schools. The Supreme Court ruled that schools and their representatives could not lead organized prayers in public schools. However, it does not prohibit silent, private prayer in schools.
Between 2000 and 2010, the illegal immigrant population of Texas increased 60 percent, while California and Florida had no increase.	Romney is essentially correct on the numbers, but his strong implication that Perry is responsible is a significant stretch.	Romney said, "California and Florida have had 60 percent increase in illegal immigration." The numbers for Texas and California are broadly supported by the best estimates available. However, they are not as reliable as Romney's statement suggests.
Says under Mitt Romney, Massachusetts was 47th in job creation.	The job-creation ranking for Massachusetts is right; the implication that Romney is solely to blame for the state's low standing does not hold up.	Castro said that under Romney, Massachusetts was 47th in job creation. The state ranked 47th in the nation in job creation.

Demagog – the best summaries

Claim	Justification	Automatic summary
Prosadíme novelu zákona o důchodovém pojištění, která zvýší základní výměru důchodu na 10 % průměrné mzdy.	Novela zákona o důchodovém pojištění, která zvýšila výměru důchodu na 10 % průměrné mzdy, již prošla legislativním procesem a příslušná ustanovení jsou od 1. ledna 2019 účinná.	Novela zákona o důchodovém pojištění, která zvýší základní výměru důchodu na 10 % průměrné mzdy, je účinná od 1. ledna 2019.
Vybírá to prostě nezávislá komise (města pro zavedení 5G, pozn. Demagog.cz).	Soutěž 5G pro 5 měst vyhlásilo společně Ministerstvo průmyslu a obchodu a Ministerstvo pro místní rozvoj v druhé polovině října. Složení, a tedy míra nezávislosti komise, nebyla v době analýzy výroku dohledatelná.	Soutěž 5G pro 5 měst vyhlásilo Ministerstvo průmyslu a obchodu, Ministerstvo pro místní rozvoj, Ministerstvo průmyslu a obchod
Začali jsme testovat od 23. ledna, tehdy tady ještě nebyl žádný případ.	K prvnímu dohledatelnému testování na onemocnění COVID-19 v České republice došlo 24. ledna 2020. V té době v ČR skutečně nebyl žádný případ infikování koronavirem.	V České republice 23. ledna 2020 nebyl žádný případ onemocnění COVID-19.
(...) tady vzniká společná iniciativa SPD a Pirátů na to, abychom se těmi platy zabývali.	Piráti a SPD požádali, aby mimořádná schůze kvůli projednání novely zákona o platech politiků proběhla ještě do konce roku 2018.	Piráti a SPD požádali o mimořádnou schůzi kvůli dalšímu projednání novely zákona o platech ústavních činitelů.

Demagog – the worst summaries

Claim	Justification	Automatic summary
<p>TAKÁČ, moderátor: A Evropská unie, Evropská komise to ocenila? Jste říkal. To znamená nějaká filozofická výhrada typu: jádro ne, to nepřipadá v úvahu, ta nezazněla? HAVLÍČEK: V žádném případě vůbec nezazněla. V tomto naopak zaznělo to, že je správně, že míváme k bezemisní ekonomice, respektive energetice ad 1 a ad 2 jasně zaznělo to, že jádro je bezemisním zdrojem. To, že někdo má tendenci čas od času, zejména z řad těch kritiků, to zpochybňovat, tak toto zaznělo doslova a do písmene od generální ředitelky, (...) tak to řekla přímo generální ředitelka generálního ředitelství Energo, Evropské komise.</p>	<p>Dne 20. listopadu 2019 jednala česká delegace MPO v Bruselu o budoucí podobě energetiky ČR a roli jádra v ní. Náměstek pro energetiku na MPO slova ministra Havlíčka potvrdil.</p>	<p>Česká republika mluví o výstavbě nového jaderného bloku v Dukovany.</p>
<p>Pak se stalo něco, čemu nerozumím, že paní ministryně (Maláčová, pozn. Demagog.cz) přestala chodit na výbory, začala posílat ředitele odborů, což je proti jednacímú řádu, takže jsme přerušili jednání. Nakonec tam tedy začali chodit kompetentní náměstci.</p>	<p>Zmíněné chování ministryně Maláčové je skutečně proti jednacímú řádu Poslanecké sněmovny. Ovšem popsaným způsobem se zachovala pouze jednou.</p>	<p>Ministryně Maláčová byla předsedkyní výboru pro sociální politiku. Na jednání však dorazila pouze jednou, na kterých nebyla přítomna náměstka.</p>
<p>Dobrá zpráva – Evropská rada se konečně na něčem dokázala shodnout. Špatná zpráva – v porovnání s návrhem Komise z května je rozpočet na výzkum ponížený o 14 miliard €. Pěkně prosím, škrtnat na výzkumu není šetřivé, ale prostě hloupé.</p>	<p>Evropská rada se 21. července dohodla na snížení rozpočtu na vědu a výzkum proti původnímu plánu o 25,17 mld. eur v kapitole Výzkum. K dalším škrtnům pak došlo v kapitolách Evropského obranného fondu či Evropského kosmického programu.</p>	<p>V květnu došlo k přijetí dohody na víceletém finančním rámci. Rozpočet na výzkum byl zvýšen o 13,7 mld. eur.</p>
<p>V mediálním prostoru se opakovaně řeší otázka vyhoštění diplomatů, jaké jsou důvody podle Vídeňské úmluvy o diplomatických vztazích. Je to skutečně pouze řada konkrétních důvodů, které mohou hostující stranu vést k vyhoštění diplomatů.</p>	<p>Vídeňská úmluva o diplomatických stycích umožňuje označit jakéhokoli diplomata jiného státu za personu non grata a tím fakticky ukončit jeho diplomatickou misi. Může se tak ovšem stát i bez udání důvodu.</p>	<p>V mediálním prostoru se opakovaně řeší otázka vyhoštění diplomatů.</p>

SumeCzech summaries

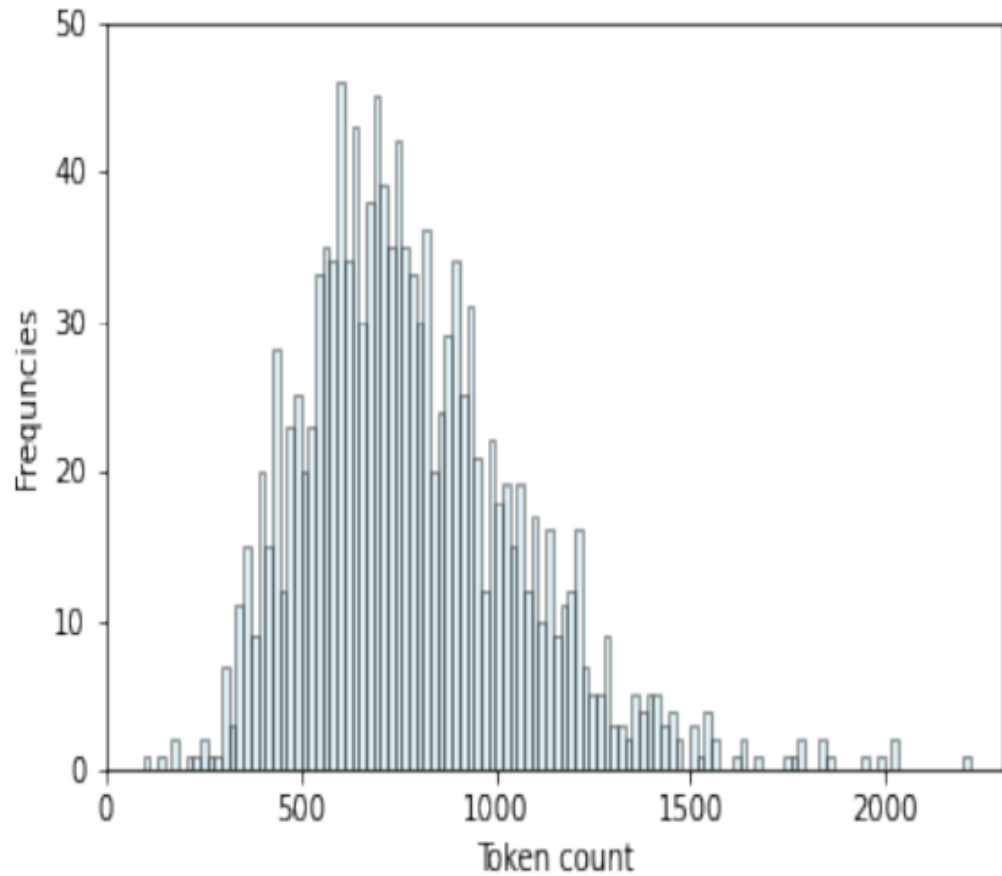
The best summaries

Headline	Automatic headline
Metal Gear Solid: Ground Zeroes	Metal Gear Solid: Ground Zeroes
Lázně Poděbrady jsou nejen na srdce	Lázně Poděbrady jsou nejen na srdce
Trains Trucks Tycoon - demo	Trains Trucks Tycoon - demo
Nejčastější zhoubné nádory u mužů	Nejčastější zhoubné nádory u mužů jsou
Rodinné domy v Unhošti poskytly sedmi rodinám vysněné bydlení	Rodinné domy v Unhošti poskytly sedmi rodinám
Neocron Arcade: The N.M.E. Project - kvalitní akční řežba	Neocron Arcade: The N.M.E. Project

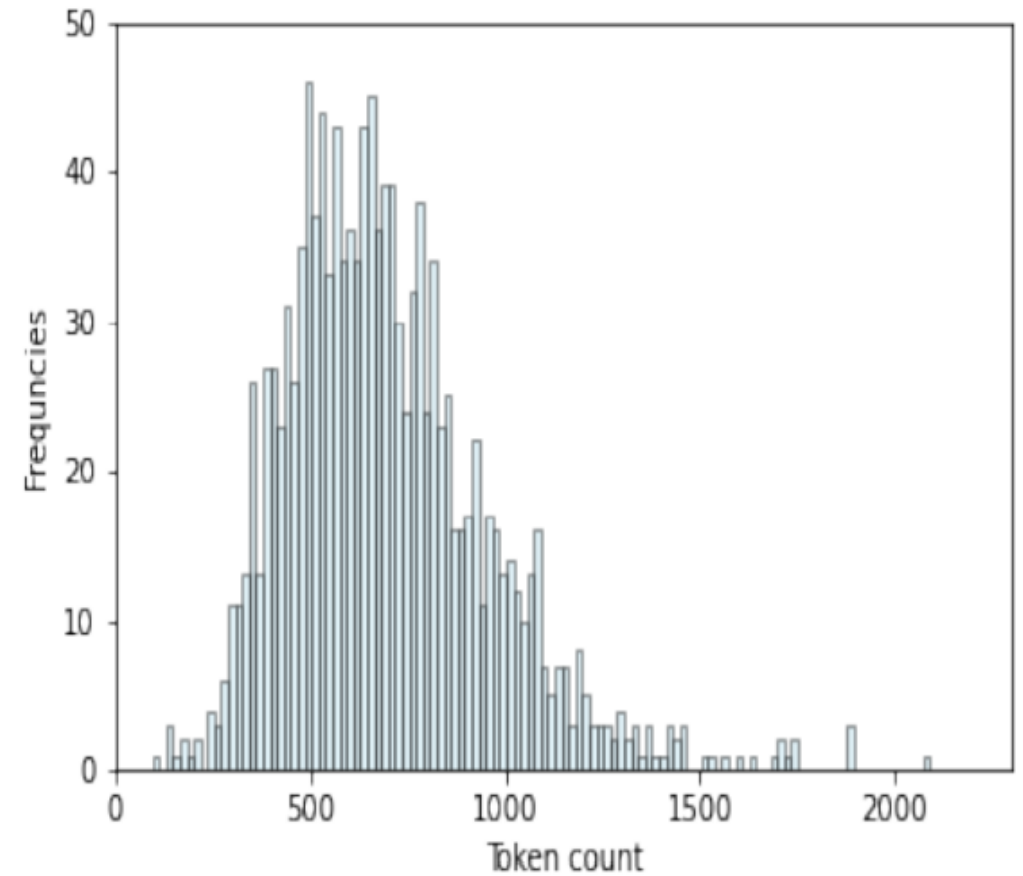
The worst summaries

Headline	Automatic headline
Makro jde blíž podnikatelům. Přejmenovává se na Vy Makro	Obchody se zajímají o malé a střed
Ojedinělá kniha o antické technologii Z východního Halabu od soboty uteklo 10 tisíc lidí	Antická trilogie Jiřího Rakušana V Sýrii se chystá poslední exodus, tvrdí
Kupte hrachovku, zvolte Juračku! Kandidát se nabízí v letácích obchodu	Poslanci chtějí ušetřit statisíce za hlas
Politici perlili: Dubové, prempos a odkloňování	Poslanci zvolili vtipného europoslance
Spolujezdec mrtvého pilota je mimo ohrožení života	Při nehodě u Korába zemřel spolujezdec

Distribution of token counts – Politifact (13 % removed)

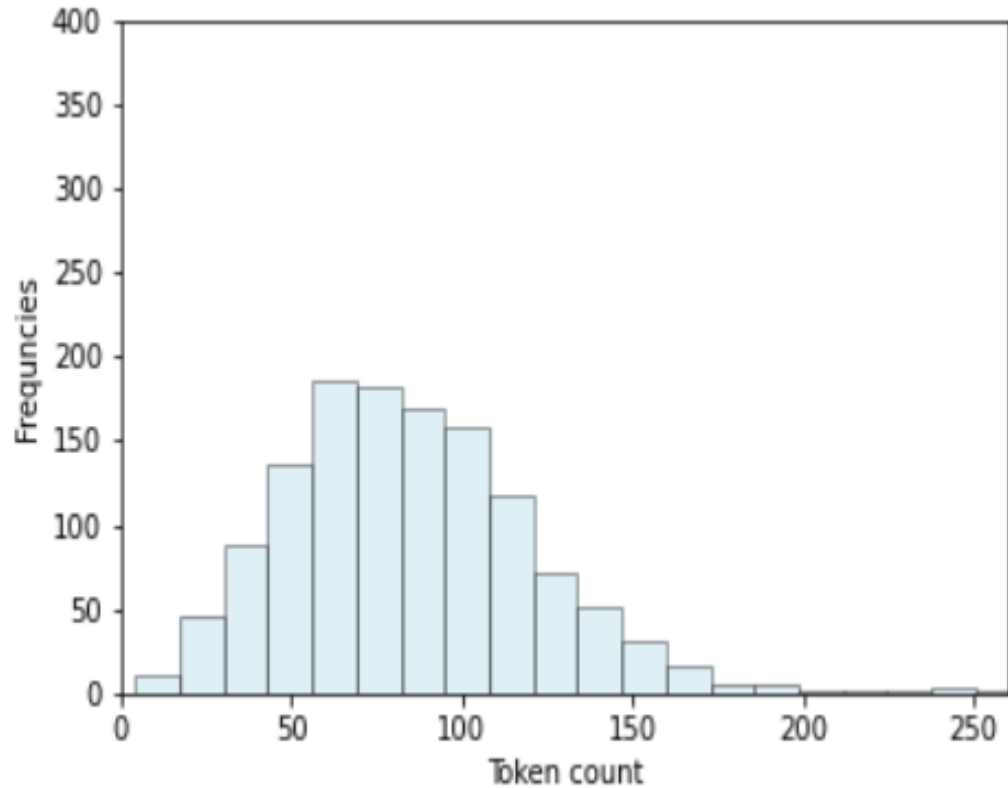


Ruling comments

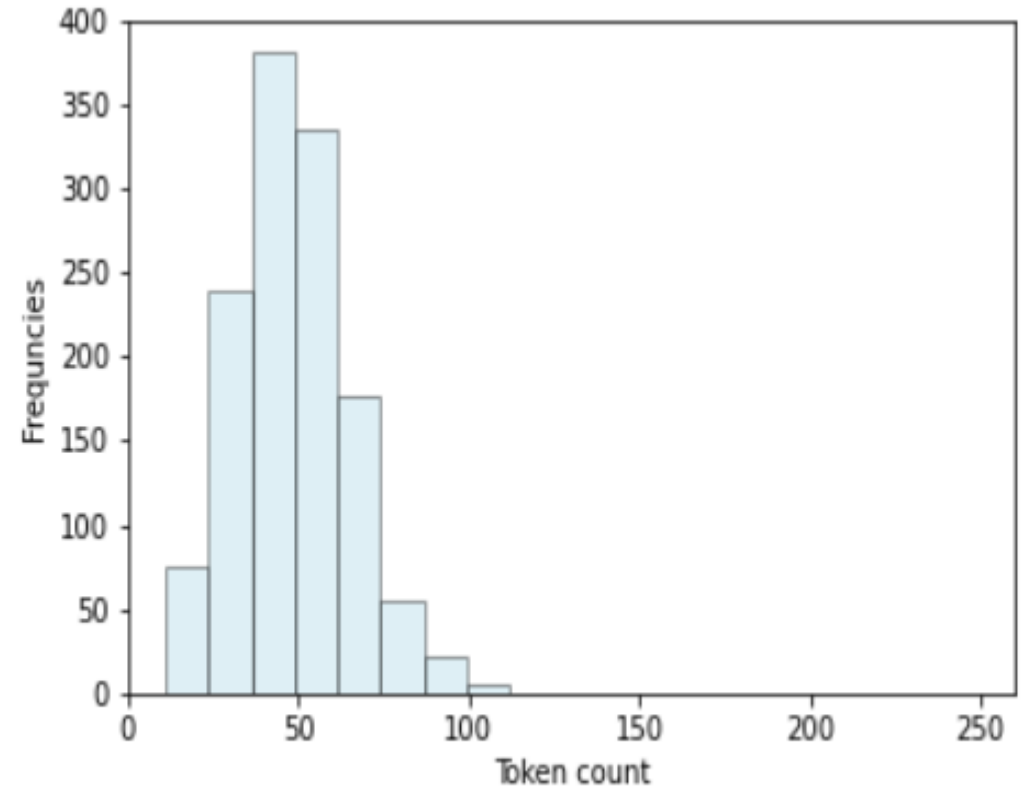


Extractive summaries of ruling comments

Distribution of token counts - Politifact

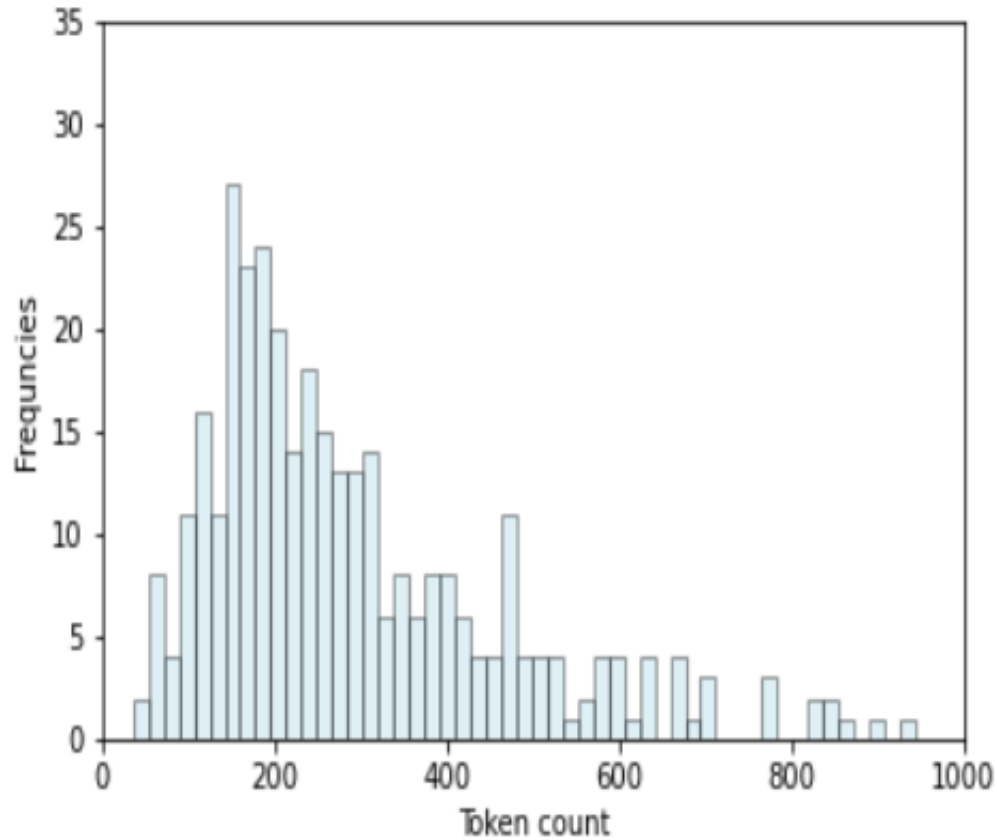


Justifications

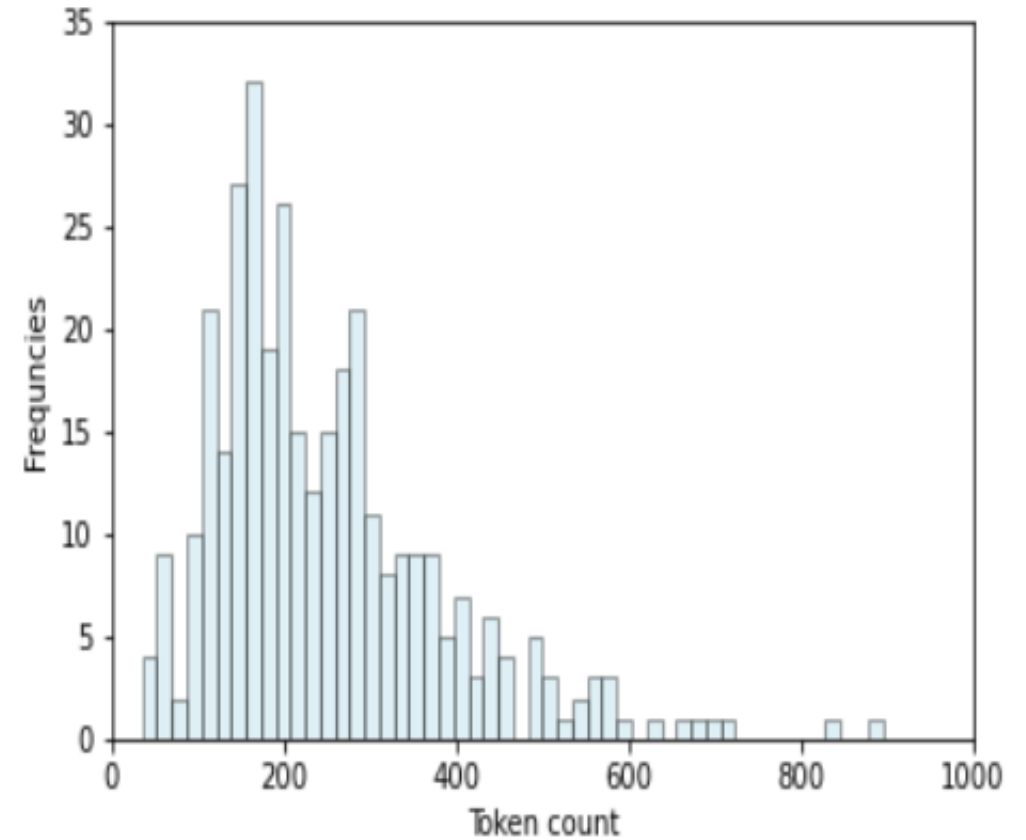


Automatic summaries

Distribution of token counts – Demagog (24 % removed)

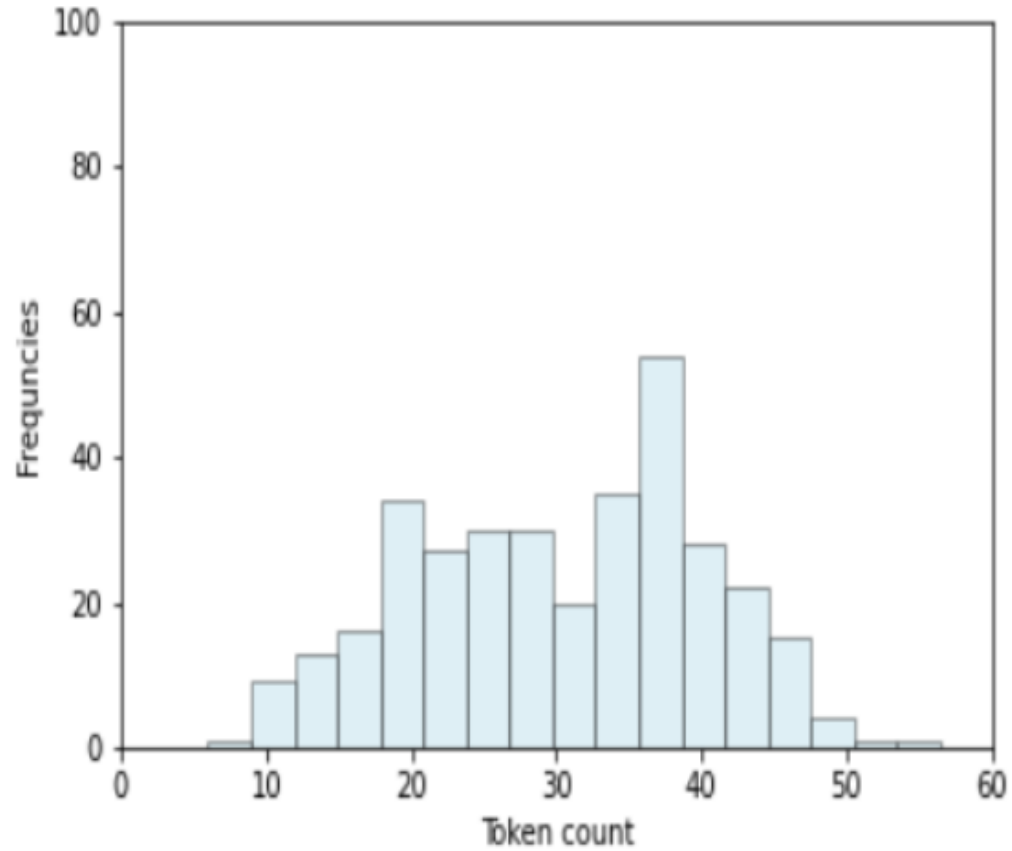


Ruling comments

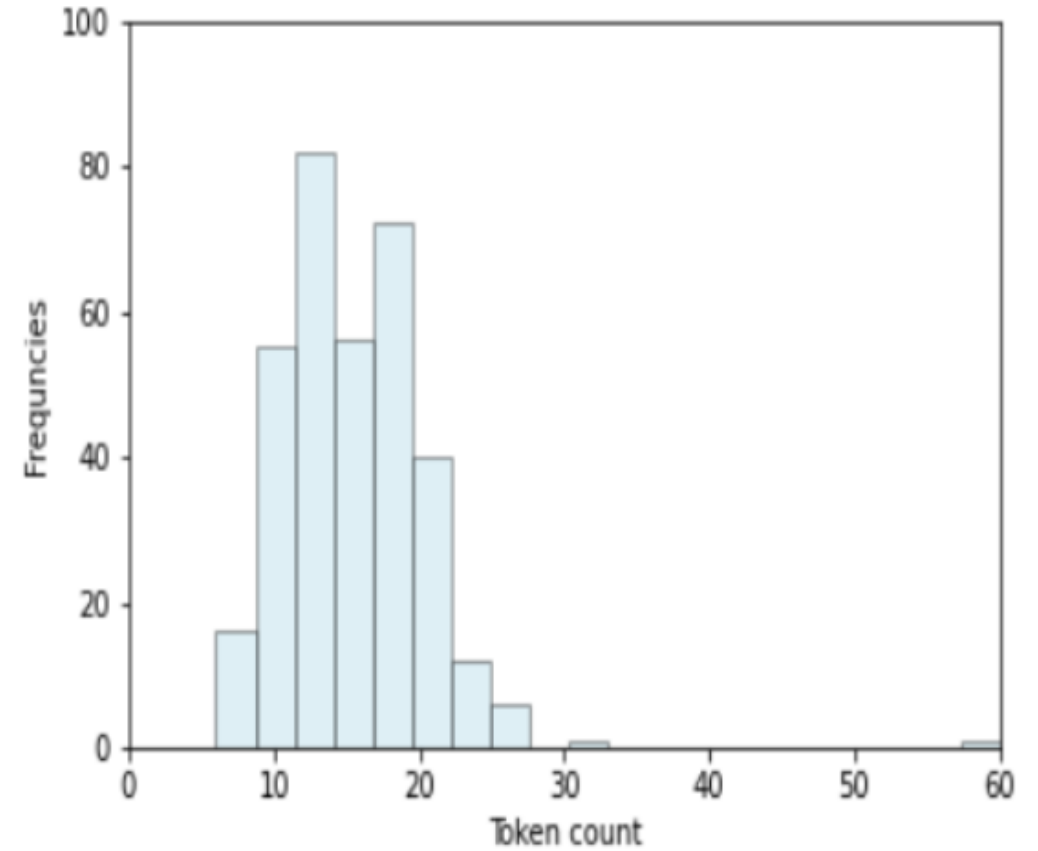


Extractive summaries of ruling comments

Distribution of token counts - Demagog

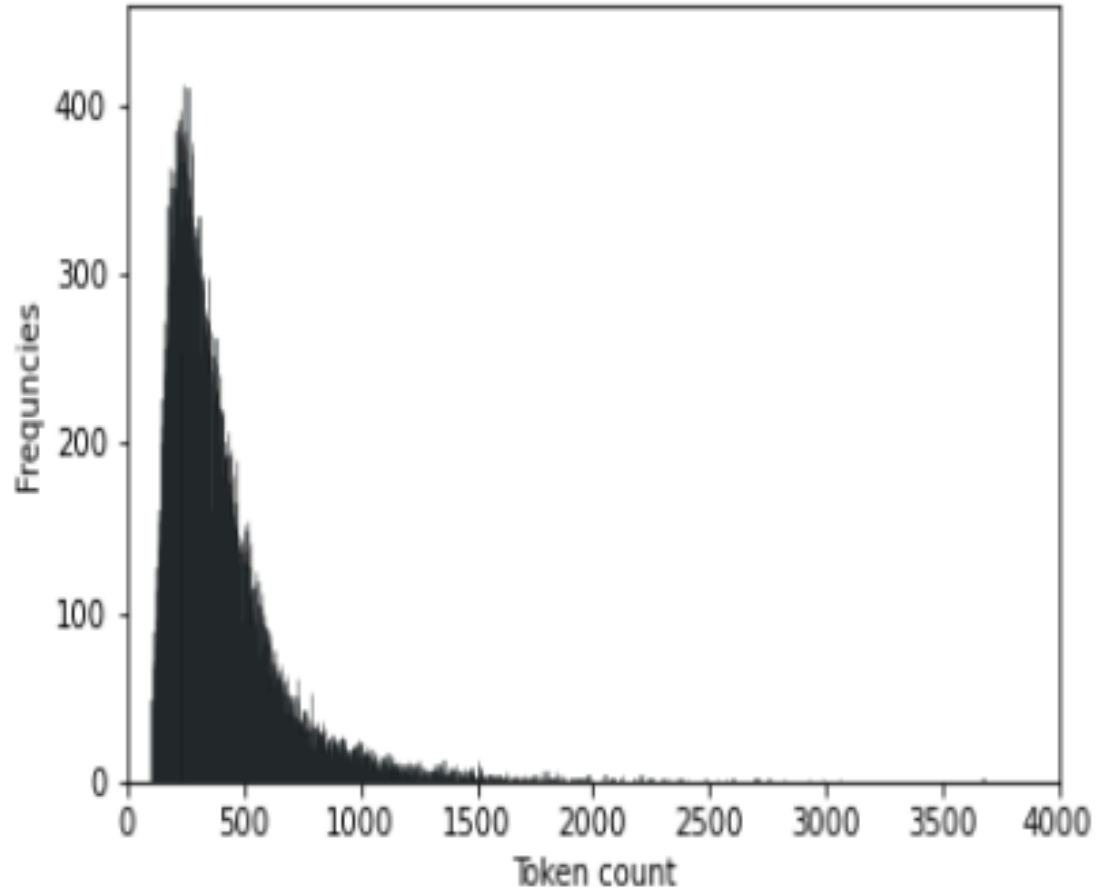


Justifications

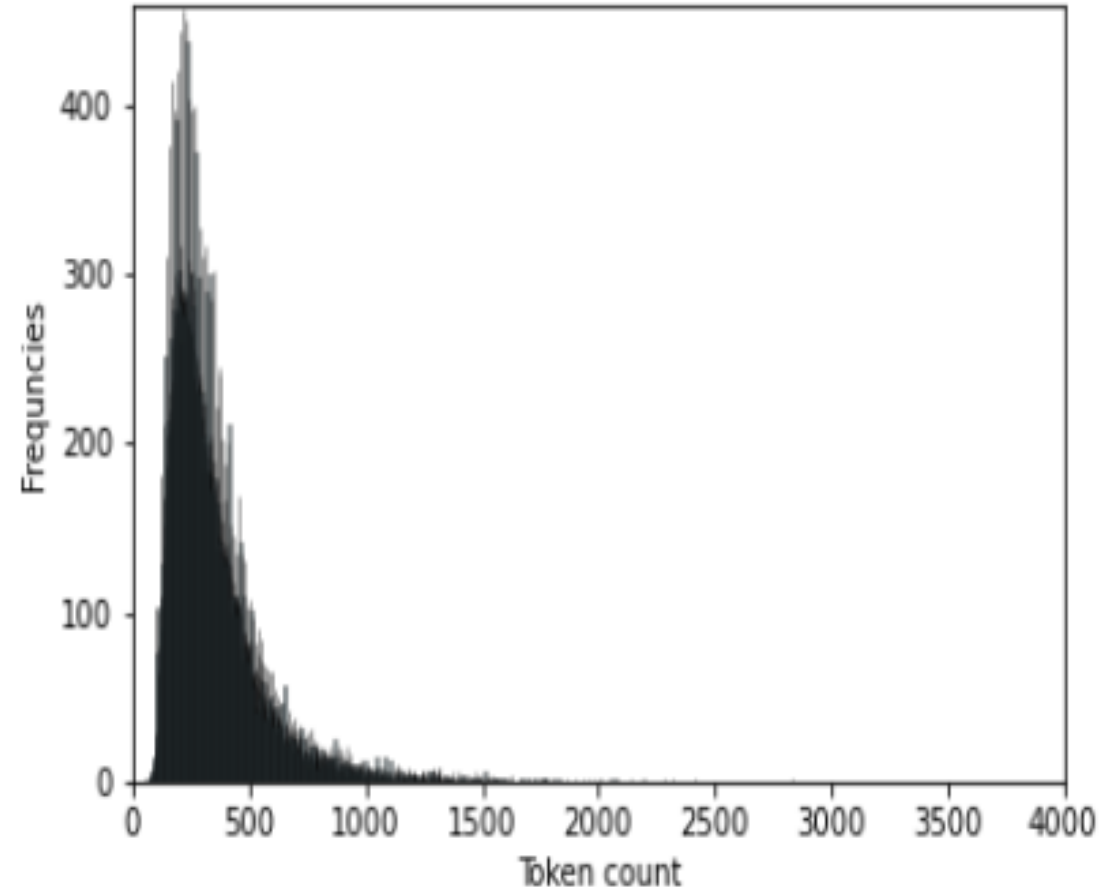


Automatic summaries

Distribution of token counts – SumeCzech (16 % removed)

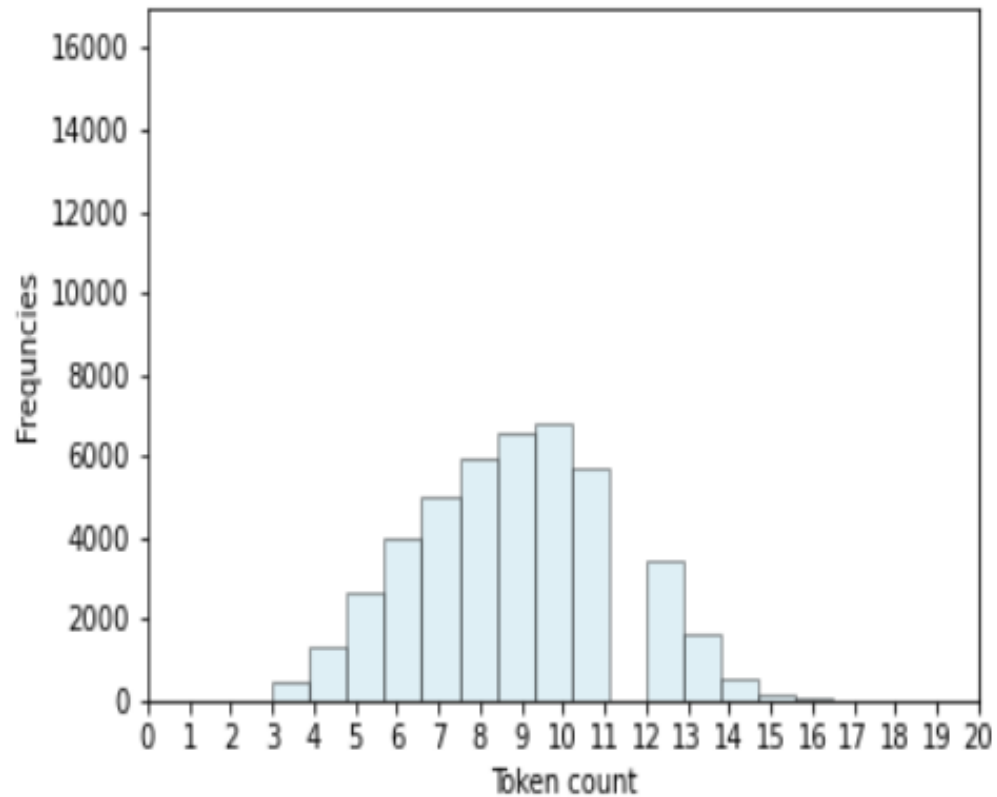


Texts

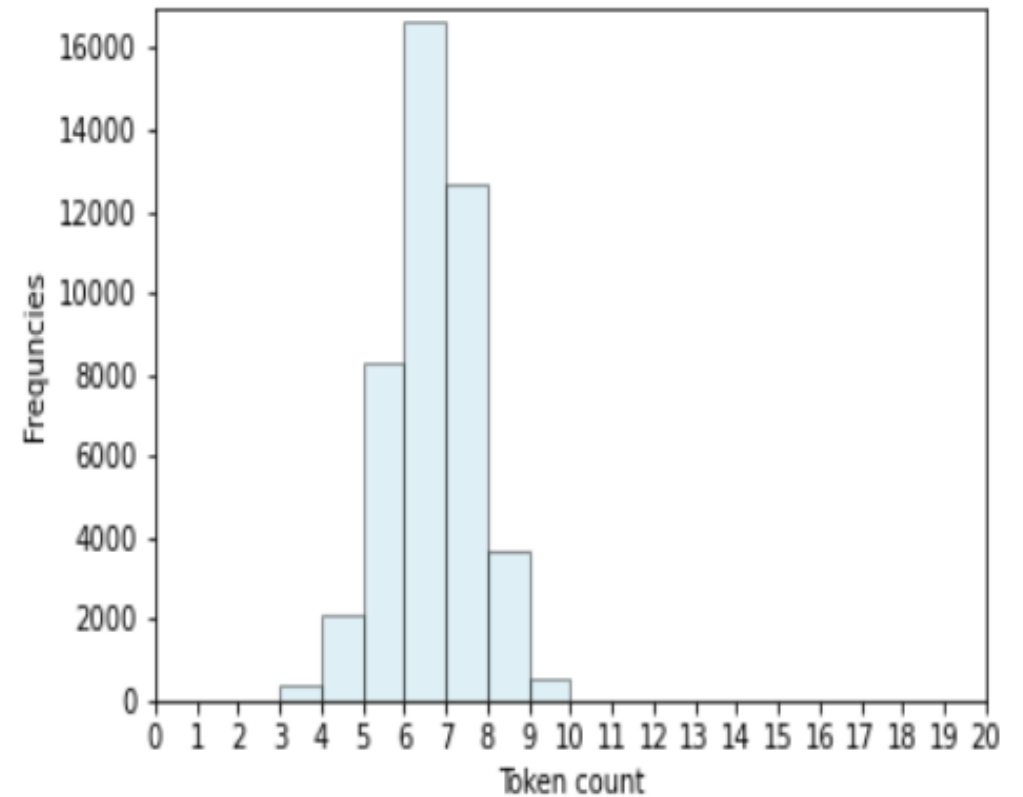


Extractive summaries of text

Distribution of token counts -SumeCzech



Headlines



Automatic short summaries

Final summary points

so far the best model for generating short summaries in the Czech language

- **improving performance of other two works**
- **only 10 % of training data used in comparison other works**

first work for automatic summarization of czech fact-checking

best results on Politifact data in comparison to other works (University of Michigan or University of Copenhagen)

quality of embeddings used for extractive summarization can affect the quality of NLG summaries

Future work:

focusing on the length of generated texts

improving the contextual representation of text

focus on the quality data selection process (possible with pre-trained models)

Thank you
for attention

