MUNI FI | NLP

Manipulative Style Recognition of Czech News Texts using Stylometric Text Analysis

Radoslav Sabol   **Aleš Horák**

xsabol@fi.muni.cz, hales@fi.muni.cz

NLP Centre, Faculty of Informatics, Masaryk University

December 10, 2022

Outline

Propaganda - what are we dealing with?

Focus on two main aspects:

- **Manipulation**
    - based on truthful events, but altered from their objective interpretation
- **Disinformation**
    - lying *intentionally*
    - doing so to *deceive* public opinion
    - commonly confused with *misinformation*, where the lying is *not intentional*
- Fake News
    - not necessarily propaganda, more focused around lying
    - satire, clickbaiting...

Misinformation?

Elton John
@eltonofficial · Follow

All my life I’ve tried to use music to bring people together. Yet it saddens me to see how misinformation is now being used to divide our world.

I’ve decided to no longer use Twitter, given their recent change in policy which will allow misinformation to flourish unchecked.

2:01 PM · Dec 9, 2022

287.7K Reply Copy link

Read 75.5K replies

Propaganda Dataset

- **benchmark dataset** of 8,644 documents
- annotation of **manipulative techniques**
  - *Argumentation, Blaming, Emotions, Demonization, Fabulation, Fear-mongering, Labelling, Relativizing*
- annotation of **document level attributes**
  - *Genre, Topic, Scope, Location, Overall Sentiment*
- annotation of **other attributes**
  - *Russia, Expert, Source, Opinion*

Example

Emotions (anger)	CS:	**Jaká xenofobie? Kdyby se nechovali jak kreténi, nikdo si jich nevšimne**
	EN:	**What xenophobia? If they didn't act like morons, no one would notice**
Fear Mongering	CS:	**Severokorejská hrozba klepe na dveře střední Evropy.**
	EN:	**The North Korean threat is knocking on the doors of central Europe.**
Labeling	CS:	Putin potvrdil novou zbraň: **nepřemožitelná jaderná hlavice**
	EN:	Putin has confirmed a new weapon: an **unstoppable nuclear warhead**
Russia (victim)	CS:	**Ať se stane cokoliv, vždy „provokuje“ Rusko.**
	EN:	**Whatever happens, Russia is the one "provoking" here.**

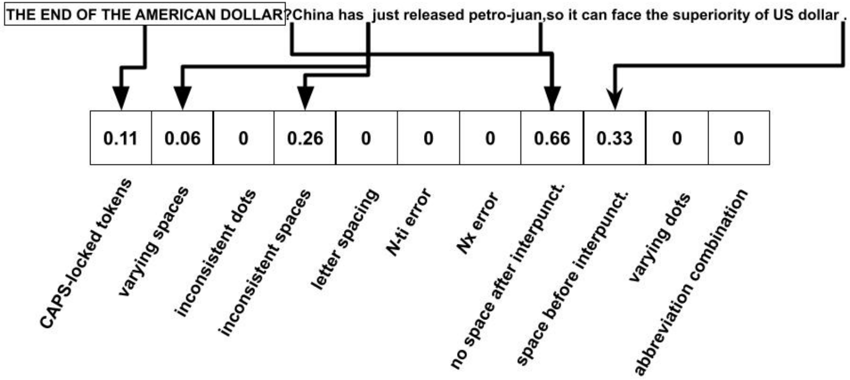
Manipulative Style Recognition

- when author intends to **deceive** and **manipulate**, his **writing style** changes
  - may be **subconscious**
- specific style can be detected via secondary features of the document - **stylometry**
- **supervised classification** of various manipulative techniques
  - techniques that *leverage emotions* or are *logical fallacies*

Proposed Stylometric Features

Feature Type	Feature Subtype	# features	Language Independent
Word Length	naive	30	✓
	improved	77	✓
	n-grams	30	✓
Sentence Length	naive	25	✓
	improved	127	✓
	n-gram	25	✓
Word Repetition	avg. repetition per sent.	1	✓
	avg. repetition per doc.	1	✓
	word class repetition	13	
	prob. word class repetition	13	
	word repetition distance	12	✓
	bag of words repetition	100	✓
Word Class N-Grams	1 to 4-grams	514	
Morphological Tags N-Grams	full	10,000	
	simplified tags	200	
Letter Casing	1 to 3-grams	77	✓
	indexed 1 to 3-grams	417	✓
Word Suffixes	stemmed	100	✓
	parametrized n-grams	325	✓
Word Richness	richness metrics	6	✓
Stopwords	for lemmas	300	✓
	for tokens	300	✓
Punctuation	frequency	11	✓
	position frequency	60	✓
	N-gram frequency	76	✓
Typography	fixed rules	11	✓
	dynamic	100	✓
Character N-Gram Distribution	1 to 5-grams	6,550	✓
Emoticons Presence	n-grams	28	✓
Total		19,529	

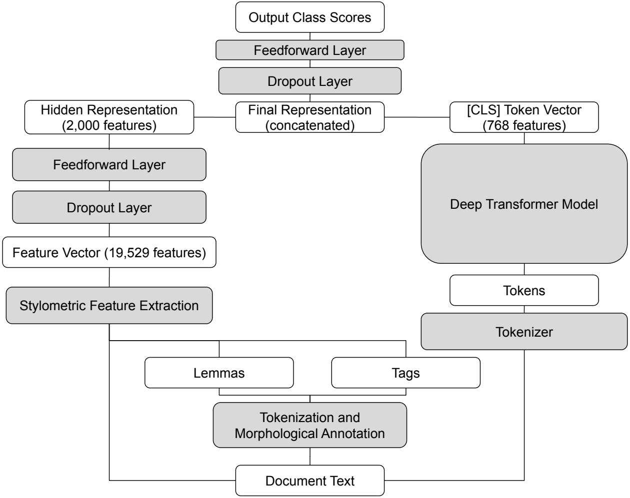
Example - Fixed Typography Rules



Attempted Approach

- **XLM Roberta Large** as the base pretrained model
- fine-tuned with the stylometric text features to enhance the existing representation
- **baseline:** majority class prediction
- non-stylometric approach with the standard classification head also present
- weighted F1 as the evaluation metric

Neural Model Architecture



Best Approaches - Manipulative Techniques

Attribute	Dummy	XLMR Large	XLMR w/ Style	Diff
Argumentation	42.46	**70.69**	70.64	-0.05
Blaming	60.67	74.55	**74.92**	0.37
Demonization	95.67	96.13	**96.19**	0.06
Emotions	77.82	81.81	**82.63**	0.82
Fabulation	74.87	80.57	**80.92**	0.35
Fear Mongering	88.89	91.71	**91.85**	0.14
Labelling	76.7	**83.37**	83.09	-0.28
Relativizing	92.27	92.75	**92.84**	0.09

Best Approaches - Document Level Properties

Attribute	Dummy	XLMR Large	XLMR w/ Style	Diff
Genre	85.99	96.46	**96.8**	0.34
Topic	10.22	**71.93**	71.12	-0.81
Scope	41.03	89.36	**90.15**	0.79
Location	20.45	82.95	**83.77**	0.82
Sentiment	74.59	**83.14**	83.06	-0.08

Best Approaches - Other Properties

Attribute	Dummy	XLMR Large	XLMR w/ Style	Diff
Expert	39.03	76.1	**77.42**	1.32
Source	44.39	52.06	**55.46**	3.4
Opinion	80.52	87.61	**88.35**	0.74
Russia	53.12	82.88	**83.63**	0.75

Conclusion and Future Work

- stylometric approaches **slightly outperform** most of the attributes
- possible direction: dealing with **domain unbalance**
  - *upsampling* does not help on massively unbalanced attributes
  - *downsampling* significantly reduces the size of already small dataset
  - text data augmentation?
- possible direction: **fine-tuning** of stylometric features
  - better feature selection
  - work towards explainability of predictions via the presented features

MUNI
FACULTY
OF INFORMATICS