Advances in Sentiment Analysis of the Large Mass-Media Documents

Nicolay Rusnachenko

nicolay-r.github.io

Newcastle University
United Kingdom

Outline

Sentiment Analysis

Sentiment Analysis

Text classification

The first attempt to propose the task[1]:

$$\langle d \rangle \rightarrow c$$

d – document
c – related class positive, negative

"The picture quality of this camera at night time is amazing"

$$\langle d \rangle \rightarrow \textit{positive}$$

[1] Peter Turney. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. 2002, pp. 417–424.

Targeted sentiment analysis

Considering entity as an input parameter[2]:

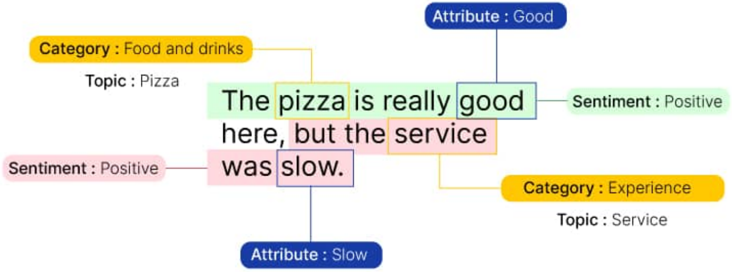
$$\langle d, e_j \rangle \rightarrow c$$

e_j – object, or entity

"The picture quality of this camera$_e$
at night time is amazing, especially with tripod$_e$"

$$\langle d, camera \rangle \rightarrow positive \qquad \langle d, tripod \rangle \rightarrow ?$$

[2] Long Jiang et al. "Target-dependent twitter sentiment classification". In: *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*. 2011, pp. 151–160.


Attribute : Good
Category : Food and drinks
Topic : Pizza
The pizza is really good here, but the service was slow.
Sentiment : Positive
Sentiment : Positive
Category : Experience
Topic : Service
Attribute : Slow

Aspect Based Sentiment Analysis

Focusing on two core tasks[3]:

1. Aspect extraction;
2. Aspect sentiment analysis:

$$\langle d, e_j, a_k \rangle \rightarrow c$$

a_k – aspect, object characteristics

"The picture quality of this camera$_e$ is amazing ..."[3]

$$\langle d, camera, picture\ quality \rangle \rightarrow positive$$

[3] Bing Liu and Lei Zhang. "A survey of opinion mining and sentiment analysis". In: *Mining text data*. Springer, 2012, pp. 415–463.

Attitude Definition

Opinions between mentioned named entities (e_j, e_m):

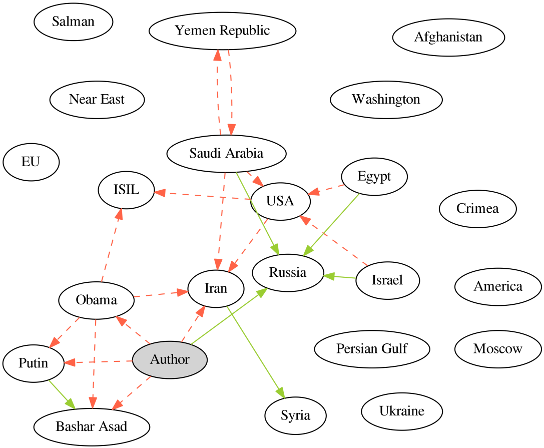
$$\langle d, e_m, e_j, a_k \rangle \rightarrow c$$

a_k – aspect
e_m – subject
e_j – object

" ... Moscow_e dissatisfied with the Warsaw's_e decision ... "

$$\langle d, e_m, e_j \rangle \rightarrow \text{neg}$$

Document-Level Attitude Representation



Evolution of Models

Approach in Large Document Sentiment Analysis

Contexts as the main idea[1]

- Retrieval of attitudes – pos and neg labeling among a set *neutrally labeled* contexts

Prediction:

- Structured output: Text Classification
- Non-structured output: Text Generation

1 Assumption: a relatively short distance between entities in the text

Rule-Based Annotation

Patterns for classification:

- Emoticons[1], matching words or phrases.

Any algorithm which allows you to perform this annotation.

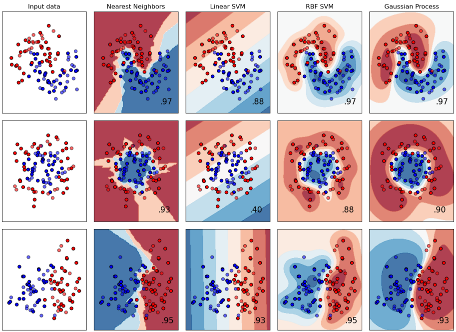
PROS: fast[2], minimal amount of `RAM` to launch
CONS: data dynamics

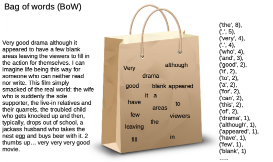
2 Relatively in comparison with further methods

Conventional Classifiers

- Documents as vectors
- NB, SVM, Random Forest, kNN.
- We can adopt different **kernels** (for the non-linear transformations)
- Every word has a scalar value: `Bag-Of-Words`

PROS: all text as vector, update.
CONS: no connection between words, vectors sparsity





Neural Networks (NN) (I)

Words as vectors, or *embeddings*:

- `One-hot` vector model

$$[0 \cdots 0, 1, 0 \cdots 0]$$

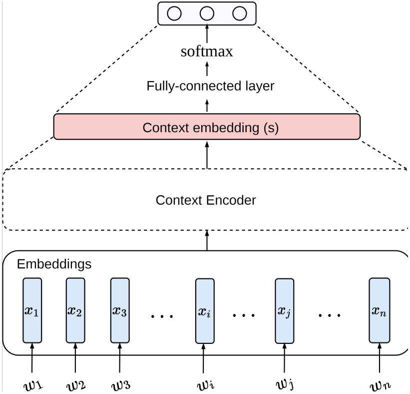
Classification: $o = W \cdot s + b$

Views of input:

- Windowed (Convolutional NN)
- Sequential (Recurrent NN)

PROS: non-linear transformations

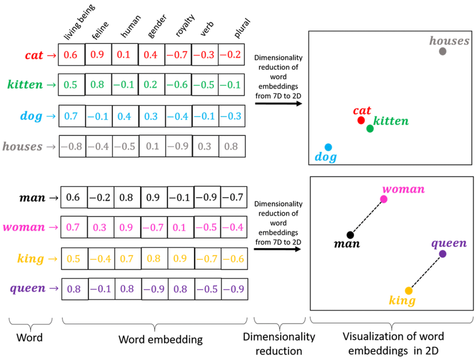
CONS: How to establish connection?



Embeddings

Raw documents could be a source of words in contexts[4]

PROS: attempt of domain/general knowledge sharing for AI models, replacement of BoW
CONS: time and resources for training on large data



[d] Tomas Mikolov et al. "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781* (2013).

Neural Networks with Embeddings

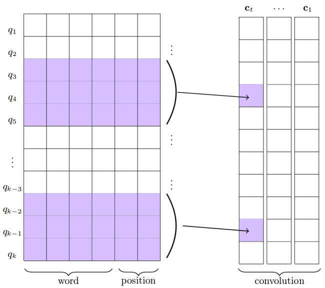


Figure: CNN, Convolution

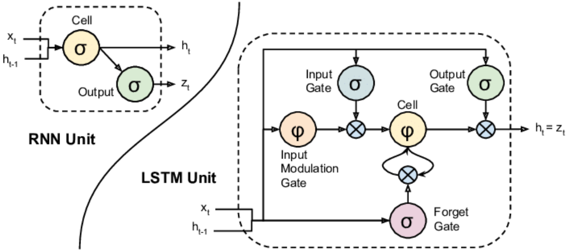
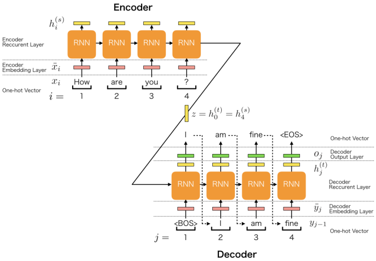


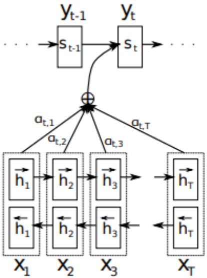
Figure: RNN/LSTM Cell

CONS: limit of window, forgetting information, limit of input in words/tokens

Attention mechanism for Machine Translation (MT)

Mechanism for assessing weights of input information, originally for MT[5]

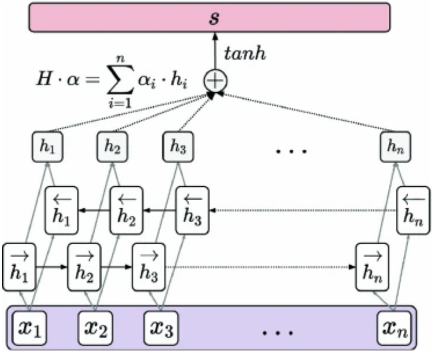




PROS: widely distributed in other NLP domains, including sentiment analysis

[5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". In: *arXiv preprint arXiv:1409.0473* (2014).

Attention for Text Classification[6]



$$m_i = \tanh(h_i)$$

$$u_i = m_i^T \cdot w$$

trainable vector

Figure: Weight calculation

[6] Nicolay Rusnachenko and Natalia Loukachevitch. "Studying Attention Models in Sentiment Attitude Extraction Task". In: *Proceedings of the 25th International Conference on Natural Language and Information Systems*. 2020. url: https://doi.org/10.1007/978-3-030-51310-8_15.

Self-Attention

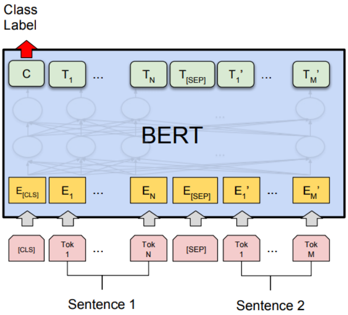
Proposed for the Machine Translation problem[7]

PROS: Affect on other NLP tasks with different conception of models training, knowledge about language
CONS: Computation cost $O(N^2)$, where N is an input sequence length

[7] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

BERT for text classification[8]

- Pre-training on large amount of data gives us a deep generalized understanding of the language, or **language model**.
- **Text classification**: FC-layer application towards the averaged embedded vectors
- Variations: RoBERTa, DistilBERT



PROS: Backbone with general knowledge

CONS: Input limitation of 512 tokens

[8] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

Long-Range Input Transformers

Main limitation for input $X \in \mathbb{R}^N$:

- $O(N^2)$ original self-attention[7] computation complexity;

Solution:

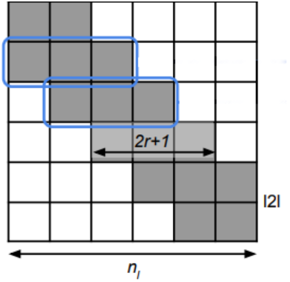
1. **Sparse self-attention**[9][10]: ETC, Longformer
2. #1 with Global Attention

[9] Iz Beltagy, Matthew E Peters, and Arman Cohan. "Longformer: The long-document transformer". In: *arXiv preprint arXiv:2004.05150* (2020).
[10] Joshua Ainslie et al. "ETC: Encoding long and structured inputs in transformers". In: *arXiv preprint arXiv:2004.08483* (2020).

Relative Position Encoding

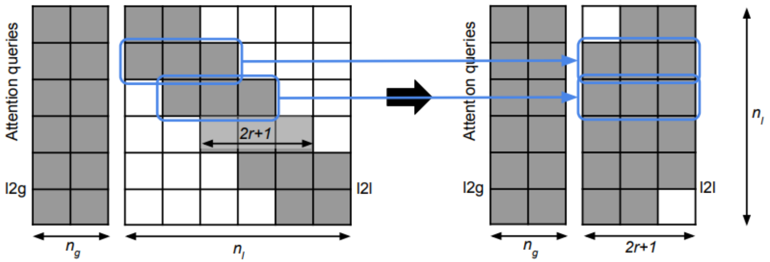
BERT[8] exploits absolute position encoding $X \in \mathbb{R}^N$. ETC proposes **relative**:

- Now position is label $l_{i,j}$ of **connection** of $x_i \in X$ with other X
- **Distance clipping**: k – limit window
  - l_k outside after i,
  - l_{-k} outside radius k before i.
- **Result** in a_l^K – learnable vectors of relative positions

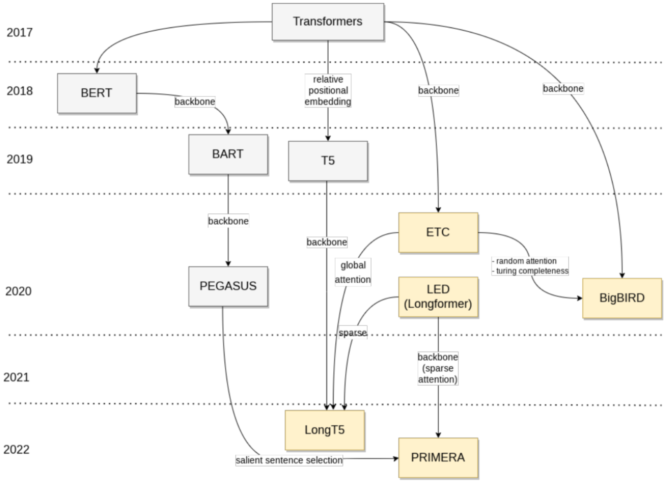


Local Attention + Global[10]

- n_l – main input components: **now windowed** (sparsed)
- n_g – global input components ($n_g << n_l$)



Evolution of the Long-Input Transformers



Data and Finetunning Advances

Supervised Learning

By default for AI methods, we consider a training based on manually annotated data by experts

PROS: Correct annotated data
CONS: Few samples, low resource domain

Supervised Learning Experiments

Trump_e accused China_e and Russia_e of “playing devaluation of currencies”

$$(\text{Trump}_{\text{subj}}, \text{China}_{\text{obj}}) \rightarrow \text{negative}$$
$$(\text{Trump}_{\text{subj}}, \text{Russia}_{\text{obj}}) \rightarrow \text{negative}$$

Supervised Learning Experiments

RuSentRel[3]: articles about Russia's international relations

Documents	**73**
Sentences per document	105.8
Entities per document	247
pos and neg pairs per document	11.47

3 https://github.com/nicolay-r/RuSentRel/tree/v1.1

RuSentRel[12] Supervised Learning Results, 3-fold cv

Model	$F_1(P, N)$
SentRuBERT	**33.4**
AttPCNN$_{\text{ends}}$	29.9
PCNN	29.6
Experts agreement	55.0

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

For `MPQA-3.0`, $F_1 = 36.0$[11]

[11] Eunsol Choi et al. "Document-level sentiment inference with social, faction, and discourse context". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016, pp. 333–343.
[12] Nicolay Rusnachenko. "Language Models Application in Sentiment Attitude Extraction Task". Russian. In: *Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS), vol.33*. 3. 2021, pp. 199–222.

Unsupervised Learning / Distant Supervision

Using external knowledge with rule-based or AI pre-trained methods to perform annotation.

emotion	*emojis*
anger	😒
joy	😂😍❤️😊💕😁☺️😀😘🙏
sadness	😭😩😢
surprise	😱

PROS: Quick data annotation for further fine-tuning
CONS: Noisy labeling

Distant Supervision Application

Main assumption: **news title** has a simple structure.

$$\ldots \quad \boxed{\textbf{Subject}_e} \quad \ldots \quad \{\textit{frame}_{A0 \rightarrow A1}\}_k \quad \ldots \quad \boxed{\textbf{Object}_e} \quad \ldots$$

Distant supervision performed in two steps[13]:

1. Collect the list A of the *most-sentiment attitudes* (*subject* $\rightarrow$ *object*) from news titles using frame $\texttt{A0}{\rightarrow}\texttt{A1}$ polarity across all news titles
2. Filter news titles and sentences, which contains at least one pair with $\texttt{A0}{\rightarrow}\texttt{A1}$ score as in A

[13] Nicolay Rusnachenko, Natalia Loukachevitch, and Elena Tutubalina. "Distant supervision for sentiment attitude extraction". In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*. 2019, pp. 1022–1030.

Datasets

RuAttitudes – automatically marked up collection of texts using the Distant Supervision approach over a large amount of mass-media short news per 2017 year.

Documents	**134442**
Attitudes per document	2.26

RuSentRel[12] distant-supervision results, 3-fold cv

Model	$F_1(P, N)$
SentRuBERT (pretrain + ft)	**37.9**
$\text{AttPCNN}_{\text{ends}}$	32.2
SentRuBERT	**33.4**
$\text{AttPCNN}_{\text{ends}}$	29.9
PCNN	29.6
Experts agreement	55.0

Prompts, prompts, prompts!

Provide additional information that mimicking the expected class or region of text to consider.

- Predefined template: QA, NLI
- Sequence of words mimicking the class[14]
- With abstract tokens serializing a particular task[15]

[14] Taylor Shin et al. "Autoprompt: Eliciting knowledge from language models with automatically generated prompts". In: *arXiv preprint arXiv:2010.15980* (2020).
[15] Xiang Lisa Li and Percy Liang. "Prefix-tuning: Optimizing continuous prompts for generation". In: *arXiv preprint arXiv:2101.00190* (2021).

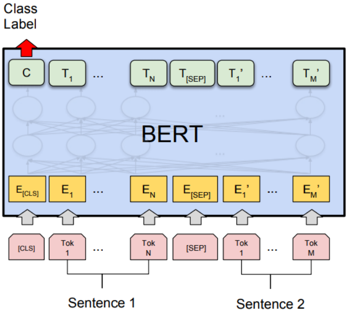
BERT with prompts[16]

Input sequences:

- TextA: Input context terms
- TextB: (Optional), **as prompt**:

$\underline{E}_{subj}$ towards $\underline{E}_{obj}$ in « $\underline{E}_{subj}$ … $\underline{E}_{obj}$ » is NEG

Context labeling: FC-layer application towards the averaged embedded vectors



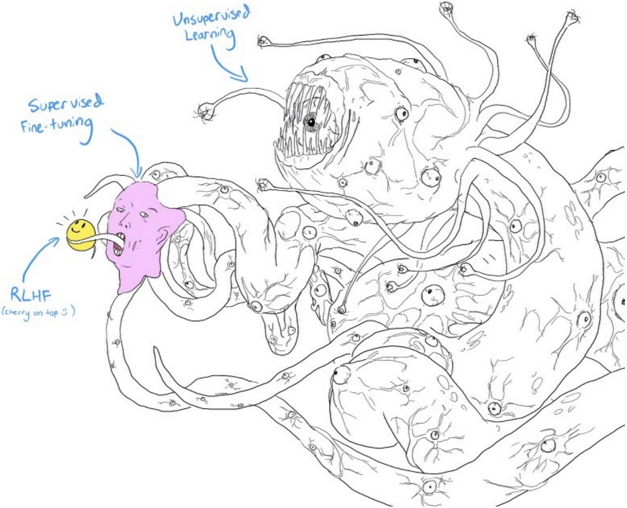
[16] Chi Sun, Luyao Huang, and Xipeng Qiu. "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence". In: *arXiv preprint arXiv:1903.09588* (2019).

RuSentRel[12] distant-supervision results, 3-fold cv

Model	$F_1(P, N)$
SentRuBERT (pretrain + ft) + $\mathrm{NLI}_{\mathrm{prompt}}$	**39.0**
SentRuBERT (pretrain + ft)	**37.9**
$\mathrm{AttPCNN}_{\mathrm{ends}}$	32.2
SentRuBERT	**33.4**
$\mathrm{AttPCNN}_{\mathrm{ends}}$	29.9
PCNN	29.6
Experts agreement	55.0

Official RuSentRel leaderboard

Large Language Models


Unsupervised Learning
Supervised Fine-tuning
RLHF
(cherry on top)

Targeted Sentiment Analysis Task[17]

RuSentNE-2023[4].

- Texts in Russian
- Texts represent sentences with mentioned objects in it.
- Volume: $\approx 6 \times 10^3$ Train, 3×10^3 Validation, 2×10^3 Test

Participant	F1-PN-macro (rank)
mtsai	66.67 (1)
cookies	66.64 (2)
lsanochkin	62.92 (3)

4 https://codalab.lisn.upsaclay.fr/competitions/9538#results

[17] Anton Golubev, Nicolay Rusnachenko, and Natalia Loukachevitch. "RuSentNE-2023: Evaluating Entity-Oriented Sentiment Analysis on Russian News Texts". In: *Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue" (arxiv:2305.17679*. 2023.

Zero-Shot[18]

Providing no information except the instruction to the model in a form of the textual **prompt**:

What is the attitude of the sentence s to the target t?
Select one from: positive, negative, neutral

PROS: No need training
CONS: Lack of context / Misunderstanding / Hallucinations

[18] Bowen Zhang, Daijun Ding, and Liwen Jing. "How would Stance Detection Techniques Evolve after the Launch of ChatGPT?". In: *arXiv preprint arXiv:2212.14548* (2022).

Translate text in English![5]

Model	Language	F1(P,N)-m
Mistral-Instruct-7B-v0.1	en	**49.56**
GPT-3.5-turbo-0613-en	en	49.36
GPT-3.5-turbo-1106-en	en	48.00
Mistral-Instruct-7B-v0.2	**ru**	46.98
GPT-3.5-turbo-0613-ru	**ru**	45.97
Mistral-Instruct-7B-v0.2	en	45.62
GPT-4-1106-preview	**ru**	45.42
Mixtral-8x-7B-Q2-offload	en	42.78
DeciLM-7B	en	42.61
FLAN-T5-xxl	en	42.47
LLaMA2-13B	en	41.39
GPT-3.5-turbo-1106	**ru**	38.88
Microsoft/Phi-2	en	37.26
Mistral-Instruct-7B-v0.1-Saiga	**ru**	37.14

5 `googletrans` Python package as an example.

Result Analysis

Example 1.

One of Sweden's most famous `actors` died in a fire.

Example 2.

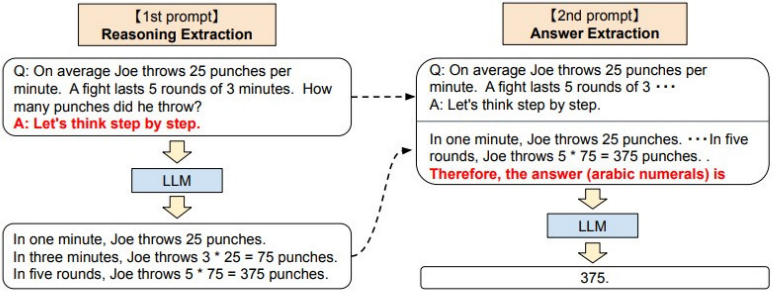
The `famous musician` lost consciousness at a concert.

Example 3.

The legendary `Chuck Berry` passed out at a concert.

Common mistakes: models classify these cases as negative, while author **sympathies** and therefore is likely has a **positive** opinion!

Zero-Shot Chain-of-Thoughts[19]



PROS: No need training

CONS: Increase of inference time

[19] Takeshi Kojima et al. *Large Language Models are Zero-Shot Reasoners*. 2023. arXiv: 2205.11916 [cs.CL].

Zero-Shot Chain-of-Thoughts

Results on RuSentNE-2023[6]:

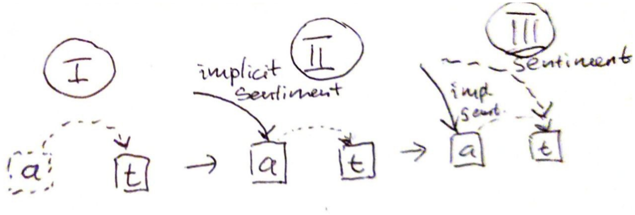
Model	F1(P,N)-m
Zero-Shot	
Mistral-Instruct-7B-v0.1	**49.56**
Flan-T5-xl	**35.35**
Zero-Shot Chain-of-Thoughts	
Mistral-Instruct-7B	46.70
FLAN-T5-xl	35.39

Conclusion: The straightaway application of the concept with *Let's think step-by-step* won't be effective in Sentiment Analysis!

6 Texts were automatically translated in English

THoR: CoT for Implicit Sentiments[20]

Three-**Ho**p **R**easoning for Sentiment:



t denotes target, while a is the aspect

[20] Fei Hao et al. "Reasoning Implicit Sentiment with Chain-of-Thought Prompting". In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2023, pp. 1171–1182.

CoT for Implicit Sentiments

Experiment: Fine tuning Flan-T5 on RuSentNE-2023 training data.
Results on RuSentNE-2023[7]:

Model	F1(P,N)-m
THoR Tuning	
Flan-T5-large	**62.29**
Flan-T5-base	59.75
Fine-Tuning	
Flan-T5-large	60.80
Flan-T5-base	57.01
Zero-shot	
Mistral-Instruct-7B-v0.1	**49.56**
GPT-3.5-turbo-0613-en	49.36
GPT-3.5-turbo-1106-en	
Mistral-Instruct-7B-v0.2	46.98
Flan-T5-xl	35.35

7 Expriemnts with texts, automatically translated in English

Conclusion

- The informative analysis of the large Mass-media texts is a granular analysis:
  - Text Classification → Targeted Sentiment Analysis → Aspect-based Analysis → Attitude Extraction
- Attitude extraction is considered as text classification problem of small text parts[8]
- The latest advances is self-attention which lead us to transformers that can memorize information from massive amount of the pretrained texts

8 Generative transformers with the largest input of 16K tokens.

Conclusion

- Rule-based
- Linear classifiers + features
- Neural Networks + embedding + attention + features
- Language Models[9] **prompting**
    - BERT / Long input Language Models
    - Large Language Models
        - Zero-Shot Language Models
        - Chain-of-Thoughts for Language Models

*The crucial part of optimizations are **prompts**[21] ... early in a form of features and later closer to output clarification*

9 Even fine-tuned Transformer classifier may outperform LLM

[21] Shuofei Qiao et al. "Reasoning with Language Model Prompting: A Survey". In: *arXiv preprint arXiv:2212.09597* (2022).

Not mentioned but Recommended

- Parameterized fine-tunning (PEFT)
- *Lost in the Middle*[10]

10 https://arxiv.org/pdf/2307.03172.pdf

Application #1: Pipelines for large collection processing

AREkit – Document level **A**ttitude and **R**elation **E**xtraction tool**kit** for sampling mass-media news into datasets for your ML-model training and evaluation

`github.com/AREkit`

Application #2: Granular View of Sentiment Relations

AREkit – granular viewer for sentiments in large texts and mass-media collections

`github.com/ARElight`

Thank you for attention!

Sentiment Attitude Extraction

Input:

1. Collection of analytical articles $\langle D_i, E_i \rangle$ (in Russian)
   - Each article includes: document D_i, list of mentioned named entities E_i
2. For synonymous mentions: given a collection of synonyms:

   Russia_e, RF_e, $\text{Russian Federation}_e$

Task: For each D_i complete the list of sentiment attitudes (pairs $\langle e_i, e_j, l_{i,j} \rangle$)[22], with label $l_{i,j} \in \{\text{pos}, \text{neg}\}$

[22] Natalia Loukachevitch and Nicolay Rusnachenko. "Extracting sentiment attitudes from analytical texts". In: *Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialogue-2018 (arXiv:1808.08932)* (2018), pp. 459–468.

Distant Supervision Experiments

1. **News collection**: Russian articles from mass-media sources (**8.8M**);
2. Knowledge Base **RuSentiFrames**[11]: describes sentiment association, conveyed by *predicate* in a form of a verb on noun (311 frames)
    - **roles**: A0 (agent), A1 (theme);
    - **dimensions**: authors attitude towards the participants mentioned in text; **polarity** – score between participants;

Frame (bragging)	Description
entries	bragging, boasting
roles	`A0`: those who bragging `A1`: the object of bragging
polarity	`A0`→`A1`, pos `author`→`A0`, neg

11 https://github.com/nicolay-r/RuSentiFrames

Example of the Distant-Supervision Technique

Title
Tillerson_e : USA_e won't remove ***sanctions***$_{neg}$ from Russia_e before the return of Crimea_e

$\downarrow$ $\texttt{USA}\rightarrow\texttt{Russia}_{neg}$, $\texttt{USA}\rightarrow\texttt{Crimea}_{neg}$

Most sentiment attitudes	
Query	Search results
$\texttt{USA}\rightarrow\texttt{Russia}_{neg}$	pair found, scores match; pos: 32%, neg: 68%
$\texttt{USA}\rightarrow\texttt{Crimea}_{neg}$	pair not found

$\downarrow$ $\texttt{USA}\rightarrow\texttt{Russia}_{neg}$

Sentence
Secretary of State **USA**$_e$ Rex Tillerson_e, speaking in Brussels_e at a meeting Foreign_e heads of NATO_e affiliates stated that the sanctions from **Russians**$_e$ will only be removed after the return of Crimea_e, according to CNN_e.

Zero-Shot

Illustrates state-of-the art results in **zero-short learning**![18]
We use the following prompt template (NLI format)[12]:

Input

What's the attitude of the sentence "[S]" from "[X]" to the target "[Y]". positive or negative.

12

ChatGPT Results

Model	$F_1(P, N)$
ChatGPT[13]	**37.7**
SentRuBERT (pretrain + ft) + $\text{NLI}_{\text{prompt}}$	**39.0**
SentRuBERT (pretrain + ft)	**37.9**
$\text{AttPCNN}_{\text{ends}}$	32.2
SentRuBERT	**33.4**
...	

Official RuSentRel leaderboard

13 We did not examine RuSentRel with the provided ChatGPT explanations