Deep Learning Frameworks and Their Evolution in Sentiment Analysis

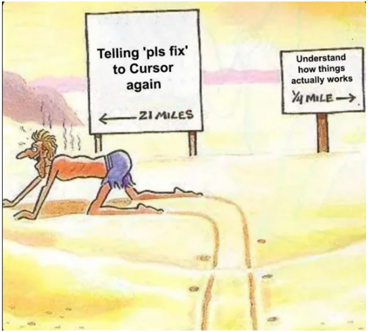
Nicolay Rusnachenko

https://nicolayr.com

Bournemouth University
CfACTs+, Center for Applied Creative Technologies
United Kingdom

Motivation

Everything is Around GenAI Nowadays ...



Answer to the following questions:

- Any concepts for proper GenAI-powered frameworks?
- Any other tasks we can relay to non-GenAI frameworks?

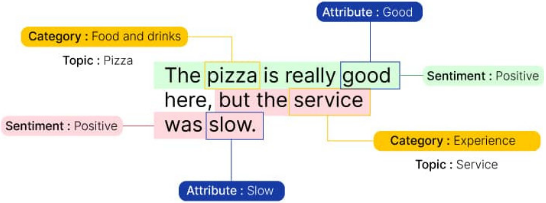
Outline

- Sentiment Analysis (Evoluation of the task)
- Deep Learning Frameworks Evolution
- Benchmarking

Sentiment Analysis Task

Origin:

Advances:



Text classification

The first attempt to propose the task[1]:

$$\langle d \rangle \to c$$

d – document
c – related class positive, negative

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

$$\langle d \rangle \to positive$$

[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$_e$"

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

[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.

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

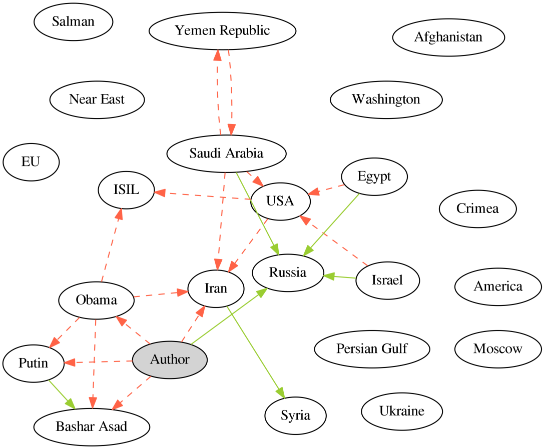
Opinions between mentioned named entities (e_j, e_m):

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

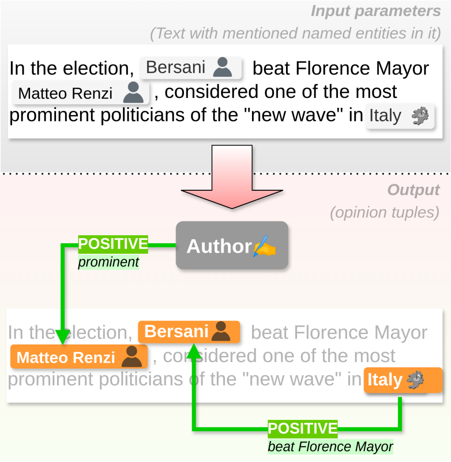
a_k – aspect
e_m – subject
e_j – object
h_t – author
t_l – time
c – sentiment class (pos, neg)

" ... Moscow$_e$ dissatisfied with the Warsaw's$_e$ decision ... "
$\langle e_m, e_j \rangle \rightarrow$ neg

Document-Level Attitude Representation



Sentiment Attitude Extraction with Explanation



Deep Learning Frameworks Evolution

Approach

Task Example: (Sentiment Analysis as Attitudes Extraction):

" ... Moscow$_e$ dissatisfied with the Warsaw's$_e$ decision ... "

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

Frameworks concept: Contexts as the main idea[1]

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

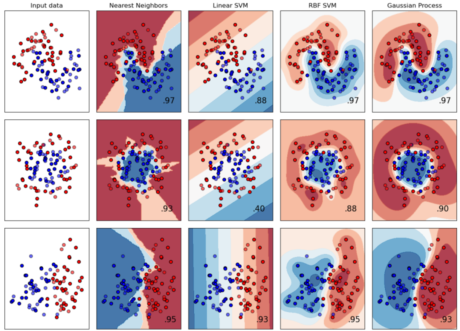
Output format:

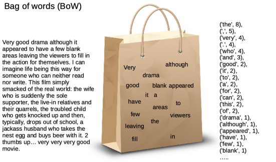
- Structured: Text Classification (before Large Language Models Era)
- Non-structured: Text Generation (Large Language Models Era)

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

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

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

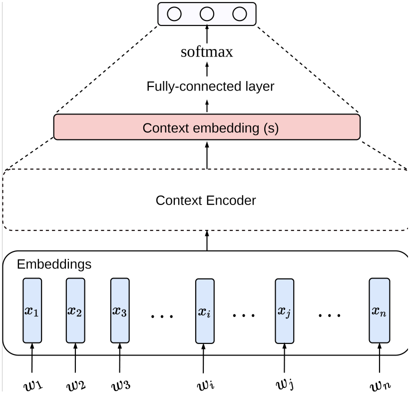
Classification: $o = W \cdot s + b$

Views of input:

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

PROS: non-linear transformations

CONS: How to establish connection?

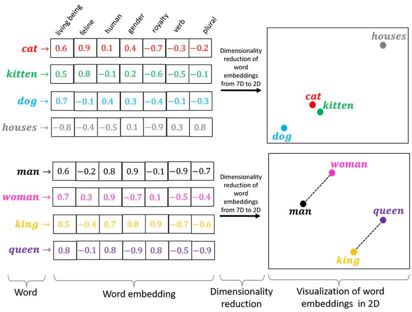


Embeddings[4]

Raw documents could be a source of words in contexts

PROS: attempt of domain/general knowledge sharing for AI models, replacement of BoW

CONS: time and resources for training on large data



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

Neural Networks with Embeddings

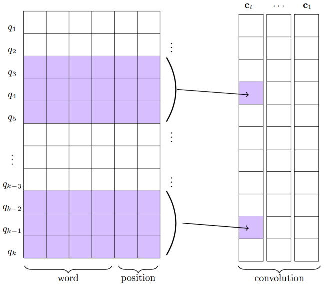


Figure: CNN, Convolution

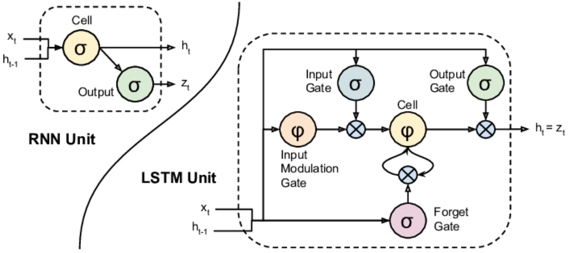
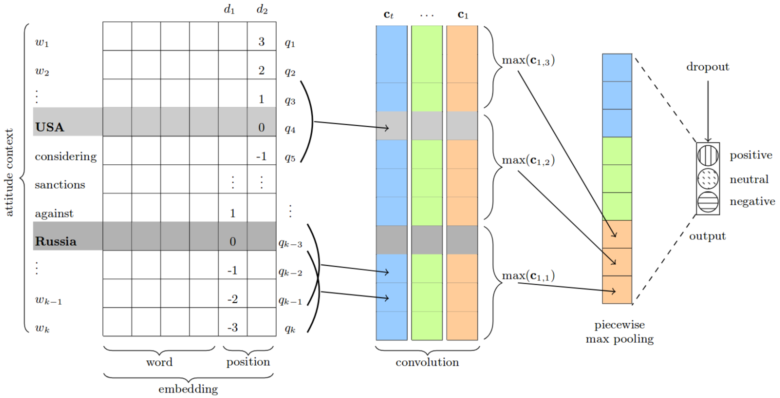


Figure: RNN/LSTM Cell

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

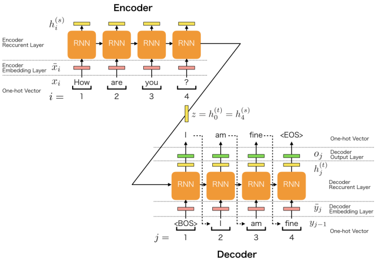
Adaptation of the Convolutional Neural Networks[5]

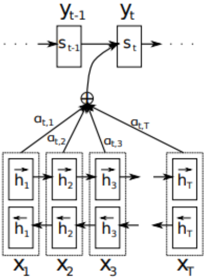


[5] Nicolay Rusnachenko and Natalia Loukachevitch. "Using convolutional neural networks for sentiment attitude extraction from analytical texts". In: *EPiC Series in Language and Linguistics* 4 (2019), pp. 1–10.

Attention mechanism for Machine Translation (MT)

Mechanism for assessing weights of input information, originally for MT[6]

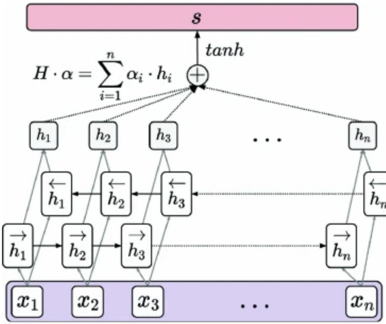




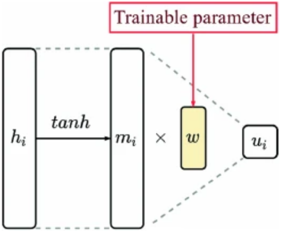
PROS: widely distributed in other NLP domains, including sentiment analysis

[6] 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[7]



(a) Context encoder architecture



(b) Quantification of h_j with respect to parameter w [17]; w represents a hidden vector which modifies during model training process

[7] 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.

Attention Visualization[8]

Att-BLSTM (Supervised Learining)

But E_{subj} consequently emphasizes its ***interest***$_{pos}$ in ***normalizing***$_{pos}$ relationships with E_{obj} (<NUM> february <NUM> year <DOT> took place the visit E at E and its ***conversation***$_{pos}$ with the spiritual leader E and with president E)

Att-BLSTM (Supervised Learining + Distant Supervision)

But E_{subj} consequently emphasizes its ***interest***$_{pos}$ in ***normalizing***$_{pos}$ relationships with E_{obj} (<NUM> february <NUM> year <DOT> took place the visit E at E and its ***conversation***$_{pos}$ with the spiritual leader E and with president E)

. . . $\textsf{Subject}_e$ . . . $\{frame_{A0 \rightarrow A1}\}_k$ . . . $\textsf{Object}_e$ . . .

[8] Nicolay Rusnachenko and Natalia Loukachevitch. "Attention-Based Neural Networks for Sentiment Attitude Extraction using Distant Supervision". In: *The 10th International Conference on Web Intelligence, Mining and Semantics (WIMS 2020), June 30-July 3, 2020, Biarritz, France.* 2020.

Advanced Attention Mechanism: «Self-Attention»

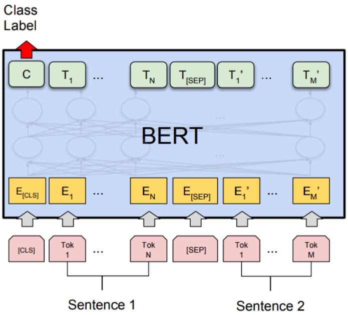
Proposed for the Machine Translation problem[9]

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

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

BERT for Text Classification[10]

- 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

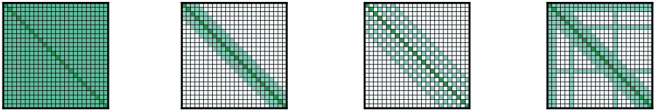


PROS: Backbone with general knowledge
CONS: Input limitation of 512 tokens

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

Encoder-Decoder / Decoder based models[12]

- Generative based: GPT, T5, Longformer, LongT5, BigBIRD
- Text classification: classification layer
- Serialized input/output[11]



(a) Full n^2 attention (b) Sliding window attention (c) Dilated sliding window (d) Global+sliding window

PROS: options to train long input with 4K, 8K, 16K

[11] Gaku Morio et al. "Hitachi at SemEval-2022 Task 10: Comparing Graph- and Seq2Seq-based Models Highlights Difficulty in Structured Sentiment Analysis". In: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics, 2022, pp. 1349–1359.

[12] Iz Beltagy, Matthew E Peters, and Arman Cohan. "Longformer: The long-document transformer". In: *arXiv preprint arXiv:2004.05150* (2020).

Domain-specific adaptation of Frameworks

Supervised Learning

Conditions when model training is based on manually annotated data by experts

Trump$_e$ accused China$_e$ of "playing devaluation of currencies"

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

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

Distant Supervision[13]

Using external **Knowledge Base (KB) rule-based** for auto-annotation.

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

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

[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.

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).

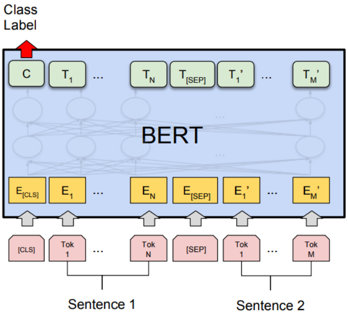
Prompt-based Tuning for Encoders (BERT)[16]

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).

Zero-shot and Few-Shot Learning for Decoders[17]

We use the following prompt template (NLI format)

Prompt

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

Format of adapting Large Language Models:

- **Zero-Shot**: No fine-tuning
- **Few-Shot**: Fine-tuning on a few examples

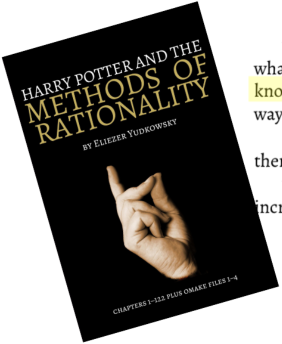
[17] Bowen Zhang, Daijun Ding, and Liwen Jing. "How would Stance Detection Techniques Evolve after the Launch of ChatGPT?". In: *arXiv preprint arXiv:2212.14548* (2022).

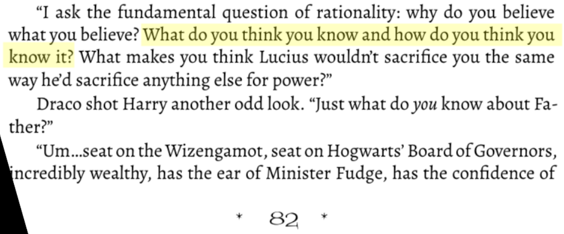
Reasoning in Sentiment Analysis

Idea: Composing a sequence of prompts (Chain of Thought)

Chain of Thought in Sentiment Analysis

Aspect → Opinion → Retrieve





"I ask the fundamental question of rationality: why do you believe what you believe? What do you think you know and how do you think you know it? What makes you think Lucius wouldn't sacrifice you the same way he'd sacrifice anything else for power?"

Draco shot Harry another odd look. "Just what do *you* know about Father?"

"Um...seat on the Wizengamot, seat on Hogwarts' Board of Governors, incredibly wealthy, has the ear of Minister Fudge, has the confidence of

* 82 *

Reasoning in Sentiment Analysis (THoR Example)[18]

THoR (Step 1): $a' = [C_1(X)$, which specific aspect of t is possibly mentioned?]

$C_1(X)$ = «Given the sentence X»

THoR (Step 2): $o' = [C_2(C_1, a')$. Based on the common sense, what is the implicit opinion towards the mentioned aspect of t, and why?]

$C_2(C_1, s')$ = «C_1. The mentioned aspect is about a'.»

THoR (Step 3): $s' = [C_3(C_2, o')$. Based on such opinion, what is the sentiment polarity towards t?]

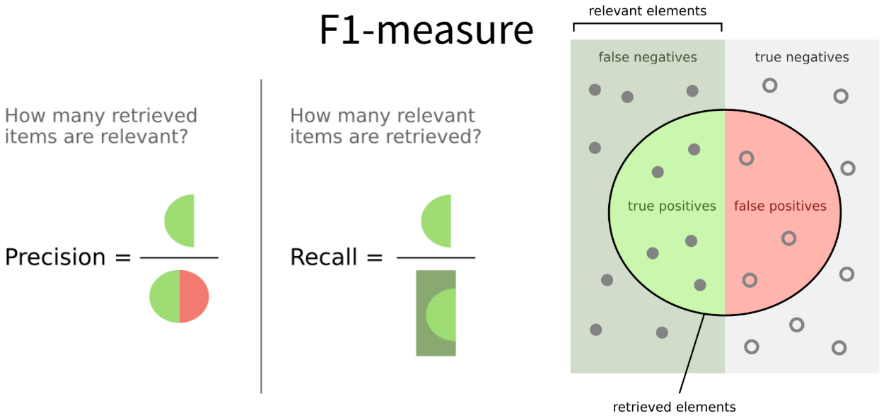
$C_3(C_2, o')$ = «C_2. The opinion towards the mentioned aspect of t is o'.»

Final label inferring: $l = [C_1$. The sentiment polarity is s'. Based on these contexts, summarize and return the sentiment polarity only, such as: positive, negative, neutral.]

[18] Natalia Loukachevitch and Natalia Tkachenko et. al. *RuOpinionNE-2024: Extraction of Opinion Tuples from Russian News Texts*. 2025.

Benchmarking

Evaluation Metric: F1 / F-Measure



Evaluation on RuSentRel dataset[19]

For the **Deep Learning Frameworks Evolution Part I**:

`https://github.com/nicolay-r/RuSentRel-Leaderboard`

[19] 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.

Evaluation on RuSentNE dataset[20]

For the **Deep Learning Frameworks Evolution Part II**:

`https://github.com/nicolay-r/RuSentNE-LLM-Benchmark`

Contributions:

- Zero-shot learning
- Reasoning

[20] Nicolay Rusnachenko, Anton Golubev, and Natalia Loukachevitch. *Large Language Models in Targeted Sentiment Analysis*. 2024. eprint: 2404.12342.

Evaluation on RuOpinionNE-2024[18]

Large Langauge Models in Few-Shot Learning:

model_name	**k=10**	**k=1**
Qwen2.5-32B-Instruct	**0.195**	**0.158**
Mistral-Nemo-Instruct-2407	<u>0.190</u>	0.112
Qwen2.5-7B-Instruct	0.184	<u>0.139</u>
Saiga-LLaMA3-8B	0.179	0.091
T-lite-it-1.0	0.157	0.096
LLaMA-3-8b-Instruct	0.153	0.119
Qwen2.5-14B-Instruct	0.145	0.121
Meta-LlaMA-3.1-8B-Instruct	0.141	0.090
RuAdapt-LLaMA3	0.123	0.073
OpenChat-3.5-0106	0.113	0.087
Qwen2.5-3B-Instruct	0.091	0.088

model_name	**k=10**	**k=1**
Qwen2.5-32B-Instruct	**0.229**	**0.204**
Mistral-Nemo-Instruct-2407	<u>0.211</u>	<u>0.157</u>
Qwen2.5-7B-Instruct	0.199	0.168
Saiga-LLaMA3-8B	0.193	0.118
LLaMA-3.1-8B-Instruct	0.173	0.110
T-lite-it-1.0	0.171	0.119
LLaMA-3-8B-Instruct	0.169	0.154
Qwen2.5-14B-Instruct	0.169	0.144
RuAdapt-LLaMA3	0.134	0.104
OpenChat-3.5-0106	0.132	0.108
Qwen2.5-3B-Instruct	0.120	0.119

Figure: Average Performance (left) and Best Performance (right)

https://arxiv.org/pdf/2504.06947

Conclusion (Deep Learning Frameworks Evolution)

- Linear classifiers + features
- Neural Networks + embedding + attention + features
- Language Models
- Language Models + **prompts**
- **Large** Language Models and Zero-Shot Learning
- + **Few-shot Learning** → Reasoning

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

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

Efficiency Tip: (It is not about prompting only)

Preprocessing: Techniques to relay LLM responsibilities

- Automated Text Translation
- Context Extraction
- Entities Masking

Thank you for attention!

https://nicolayr.com

Distant Supervision Experiments

1. **News collecion**: Russian articles from mass-media sources (**8.8M**);
2. Knowledge Base **RuSentiFrames**[2]: 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

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