



Prompt-Tuning for Targeted Sentiment Analysis in Russian

Yuliana Solomatina^(✉) and Natalia Loukachevitch

Lomonosov Moscow State University, Moscow, Russia
julianasolomatina@yandex.ru

Abstract. This paper describes prompt-based methods for targeted sentiment classification of the Russian news texts, proposed during the RuSentNE-2023 competition. This task is challenging in two following aspects: first, it is required to extract sentiment towards specific entities, not of the overall text; second, the sentiments in the news texts are often implicit, which means they are harder to detect. In the present study, several strategies of prompt-tuning BERT-like models were explored. The best result was achieved when incorporating external knowledge into the prompt verbalizer. We demonstrate that prompt-tuning methods help achieve high results while being less computationally intensive than other fine-tuning and ensembling strategies.

Keywords: targeted sentiment analysis · prompt tuning · named entities

1 Introduction

Sentiment analysis is one of the most popular areas of research in natural language processing. The task of sentiment analysis is to determine the author's attitude towards the information they convey. This application is used for various purposes, including analyzing reviews on recommendation services to automatically identify users' positive, neutral, or negative opinions; monitoring reputation of large companies; identifying users' political preferences on social networks, and many others. There are two basic approaches to solving sentiment analysis tasks: one is based on vocabularies and rules [1, 13], and the other one is based on machine learning [2]. The most effective sentiment analysis systems are those created using machine learning methods. With this approach, the data is first annotated, then significant features are identified, and finally, the most appropriate classification algorithm is chosen. Machine learning methods are much better at handling difficulties such as recognizing irony or sarcasm, resolving ambiguity in sentiment lexicon, and being oriented towards a specific domain. The best results in solving this task, as well as many other natural language processing tasks, are demonstrated by pre-trained language models based on attention mechanism [27, 28], particularly BERT [3].

Speaking of the latest approaches for classification tasks in NLP, it is important to mention the “pre-train and prompt” paradigm, which has outperformed traditional fine-tuning in many tasks. In this approach, a textual prompt is used to reformulate the downstream task in such a way that it resembles the model’s pre-training task [19, 24]. Prompt-tuning can be implemented along with fine-tuning or instead of it (the latter strategy is usually used for large-scaled models).

In the current study, we aimed to evaluate the prompt-based learning approach for targeted sentiment analysis in Russian. Models were developed and tested for predicting sentiment towards extracted named entities based on the RuSentNE dataset [20]. Pre-trained encoder ruRoBERTa [14] was used for this purpose. BERT-style models are state-of-the-art in most NLP tasks related to text data analysis and also give good results on unbalanced datasets. The experiments were based on a question-answering approach - the task was formulated as a question in natural language, which was fed into the model, and the expected output was a response that could be unambiguously compared to the class label (negative, positive, or neutral sentiment). The evaluation showed that tuning both the prompt and the model’s weights demonstrates higher scores, and more complex prompt-tuning strategies also help improve the result.

2 Related Work

Before deep learning, the task of targeted sentiment analysis was handled using systems based on vocabularies and rules, as well as classical machine learning methods. For instance, in [1], the authors proposed an algorithm based on the SentiWordNet sentiment vocabulary, and in the work [2], the algorithm proposed by the authors took into account a whole set of linguistic features. All studies were mainly conducted on the English language material, and according to the results of the SemEval-2014 competition, the best result was an F-score of 63.3 [15].

The invention of transformer models had a significant impact on research in the field of targeted sentiment analysis. Most recent studies are based on fine-tuning the BERT model. For example, in the study [5], the authors propose fine-tuning all the weights of the model for a specific task. Later, more complex approaches to targeted sentiment analysis were proposed, for example, those based on the integration of external knowledge. In [11], the authors added an additional layer of embeddings to the BERT model, encoding information from external linguistic resources (vocabularies and knowledge graphs). In [9], the authors developed this idea by adding several layers of external knowledge, achieving a state-of-the-art result for the English language (F-score 83.1). Another approach is based on the idea of prompting the model with a question to the target entity as input and obtaining an output that corresponds to the class label [6].

The next important stage in NLP came with the idea of training not only (and not necessarily) the weights of the language model, but also prompts for it

[17]. In the field of aspect-based sentiment analysis, the SentiPrompt strategy was proposed [18], the main idea of which is to embed information from linguistic resources into the text of the trained prompt.

All of the studies mentioned above were conducted on English language data. There are significantly fewer studies on Russian language material using modern approaches. For example, in [8] authors tested several deep learning models (CNN, LSTM, BiLSTM) and BERT model on Russian sentiment evaluation datasets. For the BERT-based model, the targeted sentiment analysis task was formulated as a question-answering task and as a natural-language inference task. The latter approach demonstrated the best results. In [25], the authors describe the distance supervision approach for automatic annotation of sentiment in texts, which significantly improved the classification results of neutral networks. In [23], authors proposed an attention-mechanism-based model that separately processed the sentence embedding and the target word embedding.

The prompt-tuning-based models, trained on Russian texts, is also an under-researched subject, and most existing papers describe prompt-tuning only in relation to GPT-like models while solving text generation tasks [7, 10]. In 2023, the RuSentNE-2023 competition was organized for sentiment analysis of named entities in news texts, which is expected to increase scientific interest in this area. The models described in the present paper were developed as a result of the RuSentNE-2023 competition participation. The ruRoBERTa model for classification was chosen as the baseline model during the preliminary evaluation stage of the current study. As compared with ruBERT, ruRoBERTa uses the BPE tokenizer, which helps process rare tokens more effectively. This is essential for the news data since there are many mentions of named entities. Moreover, ruRoBERTa was pre-trained on a much larger dataset, almost as large as the one used for ruGPT-3 [26].

3 Methods

The models developed in the current study were based on the prompting approach, which has been widely used for various NLP tasks and has proven to be more effective than the standard “pre-training and fine-tuning” paradigm. With this approach, a template is provided as the model’s input, which represents a piece of text formulated in such a way that the downstream task can be reduced to a language modeling task. The template consists of a *prompt*, *context*, and *target*, i.e. the language element that the model is expected to predict. This method was applied to the targeted sentiment classification in the study [6]. The authors attempted to feed the BERT model with a question about the target entity and obtain an output that corresponds to the class label. The question was formulated as follows: *What do you think of the TARGET of it?*

Despite the fact that using a manual prompt works good for downstream tasks, it is still unclear how to formulate the prompt in the most optimal way. Experiments show that even replacing one word in the template can lead to a significant change in quality metrics [19]. Moreover, we cannot claim that the

most well-formulated prompt in terms of natural language is the most optimal prompt for the model.

To maintain the advantage of using prompts while eliminating its drawbacks, the authors of the paper [19] proposed a method called *prompt tuning*. Instead of first formulating the template in natural language and then encoding them into embeddings, the authors suggest learning the prompt via gradient descent. Such prompts were called *soft prompts*, opposed to *discrete (manual) prompts*. The approach was tested on BERT and GPT and demonstrated higher quality metrics than fine-tuning.

In the current study, we applied prompt-tuning methods to the targeted sentiment analysis task. First, we tested **manual prompts** for BERT model along with some techniques to handle data imbalance (class weights and augmentation). We used the following input format:

$$[CLS]\langle sentence \rangle [SEP] \langle prompt \rangle.$$

The prompt is first initialized with manually chosen tokens. In the prompt-tuning tasks, it is also important to set a *verbalizer* - a list of words that the model should predict in the language modeling block. The predicted word is then expected to be easily mapped to a class label. For the sentiment analysis tasks, the verbalizer only usually contains class names: positive, negative, and neutral.

The example of a tunable prompt is presented below:

What:soft is the attitude:soft towards:soft X in the sentence:soft X promoted Y? [MASK]

In this example, *soft* indicates tunable tokens. Depending on a strategy, either all tokens are being tuned during training or only some of them, while others remain fixed. [MASK] indicates the masked class label, which is predicted with accordance to the verbalizer: [*positive, negative, neutral*].

We implemented several prompt-tuning strategies (they are explained in more detail in Sect. 5):

- **Mixed Template:** creating a prompt template, in which some tokens are fixed and others are tunable. They can be differentiated via special placeholders, for example, {“soft”: “What is the attitude”}{“text”: “positive, negative or neutral?”}.
- **Prompt Tuning with Rules:** learning subprompts and then aggregating them to get the class prediction [10]. The template contains several masks while the verbalizer contains a broader description. For instance, a sequence {“mask”}{“mask”}{“mask”}{“mask”} in the template corresponds to the sequence [“implicitly”, “or”, “explicitly”, “expressed”] in the verbalizer;
- **Knowledgeable Verbalizer:** extending the list of verbalizers with the words from the Russian sentiment lexicons [12] - RuSentiLex [21] and RuSentiFrames [22]. It allows the model to predict not necessarily the class name but also some word associated with it. For example, *gift, love, thankful - positive; painful, crime, harm - negative*.

4 Dataset

4.1 Dataset for the RuSentNE-2023 Competition

The RuSentNE-2023 dataset contains news texts in Russian, extracted from Wikinews. The task of the competition was to predict sentiment (positive, negative, or neutral) towards a given entity in a single sentence. The dataset was provided in the following format: *sentence*, *entity*, *entity_tag*, *entity_pos_start_rel*, *entity_pos_end_rel*, *label*. An example from the dataset is presented below.

Example 1. Sentence: After the departure of Radiy Khabirov, the administration significantly weakened, and its influence sharply decreased compared to the government apparatus,” the interlocutor told Kommersant.
Entity: Radiy Khabirov
Entity tag: PERSON
Entity pos start rel: 24
Entity pos end rel: 37
Label: 1

The dataset consisted of 12 245 pairs of sentences and entities. Table 1 presents the data distribution. It shows that the data is significantly imbalanced, which should be taken into consideration during model training.

Table 1. Data distribution in the RuSentNE-2023 dataset.

Class	Number of entries
Neutral	9580
Negative	1472
Positive	1193

The dataset was split into train, validation and test sets. The proportions of the split are presented in Table 2.

Table 2. Sizes of train, validation, and test sets for RuSentNE-2023.

Type	Size
Train	6637
Validation	2845
Test	1947

Participants of the competition could evaluate their model on the Codalab platform by uploading an archive with a file containing the model’s predictions in the format of a one-dimensional table.

4.2 Evaluation Metrics

In the current study, cross-validation was performed on three fixed splits (with a ratio of 70/30 between the train and validation sets sizes). Additionally, all models were evaluated on the split provided by the organizers of the RuSentNE-2023 competition.

The metric used for model performance evaluation was the F1-score. During cross-validation on the three splits, the F1-score was averaged across all three classes. When testing on the splits from the RuSentNE-2023 competition, the metric was averaged only across the positive and negative classes, it was the official metric of the competition. This is important for this specific task, as the main difficulty lies in extracting opinions. The neutral class, largely due to its quantitative prevalence, is classified with high precision and recall and is not of significant interest for the task at hand.

5 Experiments

5.1 Implementation Details

All the experiments were based on fine-tuning pre-trained ruRoBERTa-large model with the following hyperparameters:

- number of epochs: 5
- batch size: 16
- optimizer: AdamW
- loss function: cross-entropy

We used the largest batch size possible due to the computational limitations. The number of epochs was chosen with accordance to [9], where the original RoBERTa model was fine-tuned for a similar task.

None of the model’s weights were frozen during training, which means that both fine-tuning and prompt-tuning were implemented. The preliminary study showed that tuning only prompt tokens does not work well in this particular case, while tuning the whole model along with its prompt works better than vanilla fine-tuning. For all the prompt-tuning experiments, the OpenPrompt package was used [4].

5.2 Manual Prompts

The first block of experiments aimed to evaluate the influence of the following two factors:

- prompt type: target word or question about the target word (similar to [6]);
- class imbalance handling method (class weight calculation or augmentation).

First, the text is tokenized and padded to the maximum sequence length of 200. Then, the tokens are processed by the pre-trained model, and the [CLS] token embedding is extracted at the output, which, after regularization via a dropout layer, is fed to a fully connected layer that performs the classification.

Experiment 1. In the first experiment, the model was fed with the concatenation of the target word and the sentence. No methods to combat class imbalance were used.

Experiment 2. In the second experiment, the same data input method as in the first one was used, but weights were calculated for each class to handle data imbalance in the cross-entropy loss function (higher weights were given to smaller classes to increase loss value during training).

Experiment 3. The prompt was a question with the following formulation: *Как относятся к X?* '(How do they feel about X?), where X is the target word. No methods to combat class imbalance were used.

Experiment 4. In this experiment, the input data format is the same as in the previous one, but class weight calculation was used to handle data imbalance, as in Experiment 2.

Experiments 5, 6. In these experiments, data augmentation was implemented for the two prompt types (target word and question). To achieve greater diversity of augmented sentences, two different augmentation strategies were used. To avoid overfitting, each sentence was augmented only once. To do this, sentences belonging to the negative or positive class were extracted from the dataset, and this subset was split in half, with one half using back-translation (into English and back) and the other replacing 30% of tokens with the most similar ones based on cosine distance using the contextual embeddings of the BERT model.

5.3 Prompt-Tuning with Mixed Template

The main idea of this experiment is to use both discrete and soft prompt tokens. The template was formulated as follows:

```
{“soft”: “Какое отношение выражено к (What is the attitude towards)”}
{“placeholder”: “text_a”} {“soft”: “в предложении (in the sentence)”}
{“placeholder”: “text_b”}: {“text”: “негативное, положительное или ней-
тральное (negative, positive or neutral)”? {“soft”: “Отношение в дан-
ном предложении к (In this sentence, the attitude towards)”}{“placeholder”:
“text_a”}}{“mask”}. Here and below:
```

- soft: discrete tokens, initialized with the word embeddings and tuned during the training
- text: soft tokens
- text_a: target word
- text_b: sentence
- mask: the word that is predicted during masked language modelling

5.4 Prompt-Tuning with Rules

The idea of this experiment was to create a template where not only the verbalizer of the true class is masked (as in basic prompt tuning), but also some “auxiliary” tokens (*subprompts*) that provide a more detailed description of the task (this strategy was proposed in [16]). In general, it helps encode information, specific for the particular class. In order to evaluate how the lexical content of the prompt tokens affect the result, two templates were compared.

Template 1. {“text”: “В предложении (*In the sentence*)”} {“placeholder”: “text_b”} {“mask”} {“mask”} {“mask”} {“mask”} {“mask”} {“mask”} {“text”: “относится к (*feels about*)”} {“placeholder”: “text_a”}

The verbalizer:

```
{
0: [“автор”, “или”, “другой”, “участник”, “ситуации”, “нейтрально”],
/“author”, “or”, “another”, “participant”, “of the situation”, “neutral”/
1: [“автор”, “или”, “другой”, “участник”, “ситуации”, “отрицательно”],
/“author”, “or”, “another”, “participant”, “of the situation”, “negative”/
2: [“автор”, “или”, “другой”, “участник”, “ситуации”, “положительно”],
/“author”, “or”, “another”, “participant”, “of the situation”, “positive”/
}
```

Template 2. {“text”: “Какое отношение (*What is the attitude*)”} {“mask”} {“mask”} {“mask”} {“mask”} {“text”: “к (*towards*)”} {“placeholder”: “text_a”} {“text”: “в данном новостном тексте (*in the given news text*)”} {“placeholder”: “text_b”} {“text”: “?”} {“mask”} The verbalizer:

```
{
0: [“имплицитно”, “или”, “эксплицитно”, “выражено”, “нейтральное”],
/“implicitly”, “or”, “explicitly”, “expressed”, “neutral”/
1: [“имплицитно”, “или”, “эксплицитно”, “выражено”, “отрицательное”],
/“implicitly”, “or”, “explicitly”, “expressed”, “negative”/
2: [“имплицитно”, “или”, “эксплицитно”, “выражено”, “положительное”],
/“implicitly”, “or”, “explicitly”, “expressed”, “positive”/
}
```

Both templates contain several masked tokens, the logits of each masked token are processed into label logits, which are then combined into a single label logits of the whole task via conjunctive normal form (as proposed in [16]).

In the first template, we emphasise presence of the source of the expressed sentiment in the sentence, while the second template implies that the sentiment can be expressed either explicitly (i.e. by means of sentiment words) or implicitly (i.e. by mentioning a (dis-)advantageous fact). This presumably helps “explain” the task to the model more accurately, since each prompt points out one of the two main aspects of the targeted sentiment analysis. After tuning the prompt tokens and the model’s weights, the class label is assigned using logical rules (e.g., based on conjunctive normal form) and concatenation of the predictions of the subprompts.

5.5 Knowledgeable Verbalizer

The idea of this experiment is as follows: in the classic task setup, the model should predict the class name (“positive”, “negative” or “neutral”). However, the list of verbalizers can be extended using linguistic resources. For this purpose, Russian sentiment lexicons RuSentiFrames [22] and RuSentiLex [21] were applied.

To deal with words which are ambiguous in terms of their sentiment, we tested two strategies:

First Strategy

- RuSentiLex: if the sentiment is present in at least one of the word’s meanings, it is added to the corresponding list. If the sentiment can be both negative and positive depending on the meaning, the word is added to both lists.
- RuSentiFrames: for the predicates which are absent in RuSentiLex, check if sentiment is present in the frame and then add it to the corresponding list. If there are several possible sentiments, the word is added to both lists.

This strategy ensures high completeness for both verbalizer lists, but it allows for overlapping.

Second Strategy

- RuSentiLex: if the sentiment is present in at least one of the word senses, it is added to the corresponding list. If the sentiment can be both negative and positive depending on the meaning, the word is only added to the positive list.
- RuSentiFrames: for the predicates which are absent in RuSentiLex, check if their sentiment is present in the frame, and then add it to the corresponding list. If there are several possible sentiment, the word is only added to the positive list.

In this case, we avoid overlapping of classes and give priority to the positive sentiment class (since in all previous experiments the F1-score for this class was always the lowest). For the neutral class, the verbalizer remains unchanged. The sizes of verbalizer lists for both strategies are presented in Table 3.

Table 3. Extended verbalizer sizes.

Polarity	First strategy	Second strategy
Neutral	1	1
Positive	6350	8425
Negative	12634	12485

6 Results and Discussion

6.1 Cross-validation

The results of the experiments are provided in Table 4. In general, manually prompted models demonstrate lower metrics’ values than prompt-tuned models. Techniques aimed at handling data imbalance do not significantly improve models’ performance, although some of them can be rather time-consuming (such as augmentation). Apart from that, we assume that the perfect prompt that the model needs to perform the given task is unlikely a grammatically or semantically natural linguistic expression (the prompt formulated as a question explains the task more accurately from the human perspective, than the target word alone, but this does not lead to performance leap).

Table 4. Experimental results on cross-validation.

Prompt Type	Model	$F1_{avg}$	$F1(pos, neg)_{avg}$
Manual	target + sentence	0.651	0.522
	target + sentence + class weights	0.658	0.550
	question + sentence	0.649	0.550
	question + sentence + class weights	0.650	0.547
	target + sentence + aug	0.672	0.566
	question + sentence + aug	0.671	0.560
Soft	Mixed Template	0.687	0.580
	Prompt-tuning with rules (1)	0.727	0.630
	Prompt-tuning with rules (2)	0.697	0.635
	Knowledgeable verbalizer (1)	0.73	0.66
	Knowledgeable verbalizer (2)	0.72	0.65

When the prompt contains soft tokens, it leads to better performance (up to 0.02–0.09). Sophisticated techniques, such as prompt-tuning with rules, also contribute to an improvement (0.04–0.05 as compared with the Mixed Template strategy). However, the lexical content of the initialized prompt does not influence the final result. Since such tokens are tuned during training, many semantic aspects, though important for the natural language, are neglected.

Incorporating external knowledge into the verbalizer leads to the highest results ($F1_{avg} = 0.73$, $F1(pos, neg)_{avg} = 0.66$). It also shows that the fuller verbalizer lists have better impact on the result than the disjoint ones, while emphasising positive sentiments does not increase classification quality for this class.

After the quantitative analysis of the results, it was of particular interest to look at the predicted tokens. Generally, analyzing soft tokens doesn’t make much sense because, when projected onto natural language words, they tend to

represent a sequence of word parts or entirely non-existent words in the language. However, for the specific task, it was important to see how the predicted tokens relate to information from vocabularies.

From the model block responsible for masked language modeling (MLM), logits were extracted and then transformed into tokens. It is precisely the logits of the tokens at the output of the MLM block that are projected onto class probabilities, which means they have a substantial influence on the final prediction. After extracting these tokens, we highlighted those that are present in the vocabulary but absent in the sentence (substrings were also considered) and evaluated their impact on prediction. The statistics is demonstrated in Table 5.

Table 5. Analysis of the predictions’ statistics

Average number of vocabulary tokens in prediction	20
Maximum number of vocabulary tokens in prediction	281
Proportion of correct predictions for the neutral class given that at least 1 vocabulary token appears in the prediction	0.50
Proportion of correct predictions for the positive class given that at least 1 vocabulary token appears in the prediction	0.58
Proportion of correct predictions for the negative class given that at least 1 vocabulary token appears in the prediction	0.67
Correlation between prediction correctness and number of vocabulary tokens for positive and negative classes (using biserial criterion)	0.01 (p.value = 0.7)

Thus, it is evident that for positive and negative classes, a large proportion of correctly predicted class labels are associated with the examples in which the prediction text contains vocabulary tokens. For the negative class, this proportion is even higher than for the positive one due to the fact that its verbalizer contains more vocabulary words. Additionally, no correlation was detected between the number of vocabulary tokens and the prediction correctness.

6.2 RuSentNE-2023 Post-evaluation

As was mentioned above, the model that demonstrated the highest results both in cross-validation and the RuSentNE-2023 competition’s validation set was the model with an extended verbalizer formed according to the first strategy described in 5.5. This model was used to evaluate the quality on the RuSentNE-2023 competition’s test set during the post-evaluation stage of the competition. The results are presented in Table 6¹.

This result can be compared to the third best result of the conducted competition, lagging behind the first place by 4.51 according to the $F1(pos, neg)_{avg}$

¹ These results are as of July 2023.

Table 6. Evaluating on the RuSentNE-2023 test set

RuSentNE-2023 participants	F1 _{avg}	F1(pos, neg) _{avg}
1st place: MTS [29]	74.11	66.67
2nd place: cookies [30]	74.29	66.64
Knowledgeable verbalizer (1)	70.94	62.16

metric and by 3.17 according to the F1_{avg} metric. However, the models from the top of the leaderboard leveraged ensembling methods, which are time-consuming and computationally expensive [29, 30]. Moreover, these models were trained on additional datasets, while the model presented in the current study gained competitively high scores without dataset extension.

7 Conclusion

In this paper, we proposed a prompt-tuning method based on the knowledgeable verbalizer for targeted sentiment analysis of the Russian news texts. Specifically, we tested different learning strategies for both discrete and soft prompts as well as several technique to handle data imbalance. It turned out that prompt-tuning along with fine-tuning surpasses vanilla fine-tuning and manual prompting. Furthermore, it is relatively cheap in time and computational resources, as compared with the ensembling methods, which occupied the top of the leaderboard during the RuSentNE-2023 competition. Among prompt-tuning strategies, extending verbalizer lists via external linguistic resources performs the best, which is probably because it allows the MLM-block of the model to predict a wider range of tokens. All in all, it is quite clear that the present task, while being relatively new and under-researched, poses a challenge and presents an open field for future studies.

Acknowledgement. This work was supported by the Russian Science Foundation (grant No. 21-71-30003).

References

1. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10), Valletta, Malta. European Language Resources Association (ELRA) (2010)
2. Biber, D., Finegan, E.: Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text-Interdisciplinary J. Study Discourse* **9**, 93–124 (1989)
3. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT, pp. 4171–4186 (2019)

4. Ding, N., et al.: OpenPrompt: an open-source framework for prompt-learning. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 105–113 (2022)
5. Du, C., Sun, H., Wang, J., Qi, Q., Liao, J.: Adversarial and domain-aware BERT for cross-domain sentiment analysis. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 4019–4028 (2020)
6. Gao, Z., Feng, A., Song, X., Wu, X.: Target-dependent sentiment classification With BERT. *IEEE Access* **7**, 154290–154299 (2019)
7. Goloviznina, V., Fishcheva, I., Peskisheva, T., Kotelnikov, E.: Aspect-based argument generation in Russian. In: Proceedings of the International Conference Dialogue (2023)
8. Golubev, A., Loukachevitch, N.: Improving results on Russian sentiment datasets. In: Filchenkov, A., Kauttonen, J., Pivovarova, L. (eds.) AINL 2020. CCIS, vol. 1292, pp. 109–121. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59082-6_8
9. Hamborg, F., Donnay, K.: NewsMTSC: a dataset for (multi-)target-dependent sentiment classification in political news articles. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pp. 1663–1675, Online. Association for Computational Linguistics (2021)
10. Han, X., Zhao, W., Ding, N., Liu, Z., Sun, M.: PTR: Prompt Tuning with Rules for Text Classification. *ArXiv*, abs/2105.11259 (2021)
11. Hosseinia, M., Dragut, E., Mukherjee, A.: Stance prediction for contemporary issues: data and experiments. In: Proceedings of the Eighth International Workshop on Natural Language Processing for Social Media, pp. 32–40 (2020)
12. Hu, S., et al.: Knowledgeable prompt-tuning: incorporating knowledge into prompt verbalizer for text classification. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2225–2240 (2022)
13. Hutto, C., Gilbert, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media, 8(1), pp. 216–225 (2014). <https://doi.org/10.1609/icwsm.v8i1.14550>
14. HuggingFace ruRoBERTa-large <https://huggingface.co/ai-forever/ruRoberta-large>. Accessed 17 Jul 2023
15. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.: NRC-Canada-2014: detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 437–442 (2014)
16. Konodyuk, N.: Prompt tuning for text detoxification. In: Computational Linguistics and Intellectual Technologies, pp. 1098–1105 (2022)
17. Lester, B., Al-Rfou, R., Constant, N.: The Power of Scale for Parameter-Efficient Prompt Tuning. *ArXiv*, abs/2104.08691 (2021)
18. Li, C., et al.: SentiPrompt: Sentiment Knowledge Enhanced Prompt-Tuning for Aspect-Based Sentiment Analysis. *ArXiv*, abs/2109.08306 (2021)
19. Liu, X., et al.: GPT Understands, Too. *ArXiv*, abs/2103.10385 (2021)
20. Loukachevitch N., et al.: NEREL: A Russian Dataset with Nested Named Entities, Relations and Events. In: Proceedings of RANLP-2021, pp. 880–889 (2021)
21. Loukachevitch, N., Levchik, A.: Creating a General Russian sentiment Lexicon. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16), pp. 1171–1176 (2016)

22. Loukachevitch, N.V., Rusnachenko, N.: Sentiment frames for attitude extraction in Russian. In: Computational Linguistics and Intellectual Technologies, pp. 541–552 (2020)
23. Naumov, A., Rybka, R., Sboev, A., Selivanov, A., Gryaznov: Neural-network method for determining text author’s sentiment to an aspect specified by the named entity. In: CEUR Workshop Proceedings, pp. 134–143 (2020)
24. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8), 9 (2019)
25. Rusnachenko, N., Loukachevitch, N., Tutubalina, E.: Distant supervision for sentiment attitude extraction. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), pp. 1022–1030 (2019)
26. ruT5, ruRoBERTa, ruBERT: how we trained a series of models for Russian (*translated from Russian*)<https://habr.com/ru/companies/sberbank/articles/567776/>. Accessed 6 Sep 2023
27. Sun, C., Huang, L., Qiu, X.: Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. North American Chapter of the Association for Computational Linguistics, pp. 380–385 (2019)
28. Yin, D., Meng, T., Chang, K.: SentiBERT: a transferable transformer-based architecture for compositional sentiment semantics. In: Annual Meeting of the Association for Computational Linguistics, pp. 3695–3706 (2020)
29. Podberezko, P., Kaznacheev, A., Abdullayeva, P., Kabaev, A.: HALF-MASKED Model for Named Entity Sentiment analysis. In: Proceedings of the International Conference “Dialogue (2023)
30. Glazkova, A.: Fine-tuning text classification models for named entity oriented sentiment analysis of russian texts. In: Proceedings of the International Conference “Dialogue (2023)