Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2023"

June 14–16, 2023

# Fine-tuning Text Classification Models for Named Entity Oriented Sentiment Analysis of Russian Texts

**Anna Glazkova**
University of Tyumen
Tyumen, Russia
a.v.glazkova@utmn.ru

**Abstract**

The paper presents an approach to named entity oriented sentiment analysis of Russian news texts proposed during the RuSentNE evaluation. The approach is based on RuRoBERTa-large, a pre-trained RoBERTa model for Russian. We compared several types of entity representation in the input text, and evaluated strategies for handling class imbalance and resampling entity tags in the training set. We demonstrated that some strategies improve the results of pre-trained models obtained on the dataset presented by the organizers of the evaluation.

**Keywords:** targeted sentiment analysis, named entities, named entity oriented sentiment analysis, text classification, RuSentNE, RuRoBERTa.

**DOI:** 10.28995/2075-7182-2023-22-104-116

# Дообучение моделей классификации текстов для анализа тональности к именованным сущностям в русскоязычных текстах

**Анна Глазкова**
Тюменский государственный университет
Тюмень, Россия
a.v.glazkova@utmn.ru

**Аннотация**

В статье описывается подход к анализу тональности к именованным сущностям в новостном тексте на русском языке, предложенный в рамках соревнования RuSentNE. Подход основан на использовании RuRoBERTa-large, предобученной модели RoBERTa для русского языка. Мы сравнили эффективность нескольких типов представления именованных сущностей в тексте и оценили ряд стратегий преодоления дисбаланса классов и типов сущностей в исходном датасете. Некоторые из рассмотренных стратегий улушили качество моделей классификации текстов на текстовом корпусе, предоставленном организаторами соревнования.

**Ключевые слова:** анализ тональности к сущностям и аспектам, именованные сущности, анализ тональности к именованным сущностям, классификация текстов, RuSentNE, RuRoBERTa.

## 1 Introduction

Designing effective methods for different levels of sentiment analysis is a crucial task of natural language processing. Recently, there is a growing interest in detecting sentiment for entities instead of the whole sentence or document (Li and Lu, 2017). The task of entity-level sentiment analysis is more challenging but is more useful in many applications such as content analysis and opinion mining systems.

The paper describes a system developed for the Dialogue 2023 shared task on Targeted Sentiment Analysis for the Russian Language — RuSentNE (Golubev et al., 2023). The task aims to predict sentiment labels towards named entities in Russian news texts. In this work, we compared several pre-trained language models, types of entity representation, and strategies for processing imbalanced datasets. We found that some strategies for handling class imbalance and resampling entity tags can improve the performance of pre-trained models. Our approach based on the use of RuRoBERTa-large achieved a high result during the evaluation phase. For the final submission, we utilized a soft-voting ensemble of the models fine-tuned on the augmented dataset containing the official training set provided by the organizers, and the development set with silver labels.

The paper is organized as follows. Section 2 contains a brief review of related works. In Section 3 we describe the RuSentNE task. In Section 4 we present the methods we used. Section 5 provides and discusses the results. Some examples of the model's errors are demonstrated in Section 6. Section 7 concludes this paper.

## 2 Related Work

The problem of named entity oriented sentiment analysis relates to the field of targeted sentiment analysis. Target-based sentiment analysis involves opinion target extraction and actual target sentiment classification. Most of the existing studies usually explored one of these two sub-tasks alone (Wan et al., 2020). For example, the task of detecting the opinion target mentioned was solved using unsupervised (Yin et al., 2016; Giannakopoulos et al., 2017; Wu et al., 2018) and supervised (Xu et al., 2018; Yang et al., 2020) methods. The second sub-task, which is the target sentiment classification, aims to determine the entity-level sentiment for specific entities in each input text. In recent years numerous studies have extensively studied the target sentiment classification task. Most of the approaches were based on deep learning, including Recurrent Neural Networks (RNN) (Ye and Li, 2020), Long Short-Term Memory networks (LSTM) (Ma et al., 2018a; Ma et al., 2018b), Gated Recurrent Units (GRU) (Liu et al., 2018; Setiawan et al., 2020), and Bidirectional Encoder Representations from Transformers (BERT) (Sun et al., 2019; Wan et al., 2020; Mutlu and Özgür, 2022).

The concept of targeted sentiment analysis is relatively rarely found in works on the analysis of Russian texts. However, in recent years, a number of authors conducted research in related fields, such as aspect-based sentiment analysis and stance detection for the Russian language. In contrast to target sentiment analysis, which determines the opinion polarity towards the target entity in a given text, aspect-based sentiment analysis evaluates the polarity towards different aspects of a single entity (Saeidi et al., 2016). Stance detection aims to determine the position of a person from a piece of text towards a target (a concept, idea, event, etc.) either explicitly specified in the text or only implied (Küçük and Can, 2021). The general state of sentiment analysis research for the Russian language is reflected in (Smetanin, 2020; Loukachevitch, 2021).

SentiRuEval, the first sentiment analysis evaluation for Russian, was organized in 2015 (Loukachevitch et al., 2015). One of the tasks was the aspect-oriented analysis of the reviews about restaurants and automobiles. The participants utilized the methods based on LSTM (Tarasov, 2015), Support Vector Machines (SVM) (Ivanov et al., 2015; Mayorov et al., 2015), Conditional Random Fields (CRF) (Rubtsova and Koshelnikov, 2015), rule-based techniques (Vasilyev et al., 2015), and the use of Pointwise Mutual Information (PMI) and semantic similarity measures (Blinov and Kotelnikov, 2015). The SentiRuEval dataset was later used as a part of the official dataset during the international SemEval aspect-based sentiment evaluation (Pontiki et al., 2016) and utilized for evaluating deep-learning models. In (Kotelnikova et al., 2022), the authors compared several lexicon-based methods with RuBERT (Kuratov and Arkhipov, 2019). Within this comparison, the best result for the SentiRuEval dataset was obtained using the Russian adaptation of a Semantic Orientation CALculator (SO-CAL) (Taboada et al., 2011).

Studies in the field of aspect-based sentiment analysis on other text corpora were also carried out. In (Naumov et al., 2020), the authors presented an approach to aspect-based sentiment analysis where a named entity is considered as an aspect. The paper describes the dataset collected using a crowdsourcing platform and a deep neural model with Embeddings from Language Models (ELMo) (Peters et al., 2018) for word vector representation. The dataset for aspect-based sentiment analysis of Russian users' comments about COVID-19 was presented in (Nugamanov et al., 2021). The best result on this corpus was obtained using the RuBERT model in the Natural Language Inference (NLI) formulation. In (Makogon and Samokhin, 2022), a multilingual Ukrainian and Russian dataset for entity-oriented sentiment analysis was presented. The best result in terms of the F1-score for this dataset was obtained by RuBERT. The same model was applied for named entity oriented sentiment analysis in media texts in (Salnikova

| Characteristic | Train | Development | Test |
|---|---|---|---|
| Number of sentences | 6,637 | 2,845 | 1,947 |
| Avg number of tokens | 33.07±17.74 | 33.56±16.37 | 31.44±14.5 |
| Distribution of tags | | | |
| Country | 1,274 | 533 | 363 |
| Nationality | 276 | 116 | 110 |
| Organization | 1,487 | 653 | 484 |
| Person | 1,934 | 857 | 480 |
| Profession | 1,666 | 686 | 510 |

Table 1: The data statistics.

and Kyrychenko, 2021).

As targeted sentiment analysis involves determining the point of view of the text's author in relation to the given entity, it is related to the stance detection task. In (Vychegzhanin and Kotelnikov, 2017), several traditional machine-learning methods were evaluated on the dataset containing opinions of users about the topic of vaccinating children. Later, the dataset was complemented by the texts concerning other socially significant issues (Vychegzhanin and Kotelnikov, 2019). In (Lozhnikov et al., 2020), RuStance, a new dataset of Russian tweets and news comments from multiple sources, was presented. In 2022, the first evaluation on stance detection for Russian was organized (Kotelnikov et al., 2022). The participants analysed VKontakte users' comments discussing COVID-2019 news texts. The highest F1-score was obtained by the NLI-BERT system (Alibaeva and Loukachevitch, 2022) based on COVID-Twitter-BERT (Müller et al., 2020).

## 3 Task Description

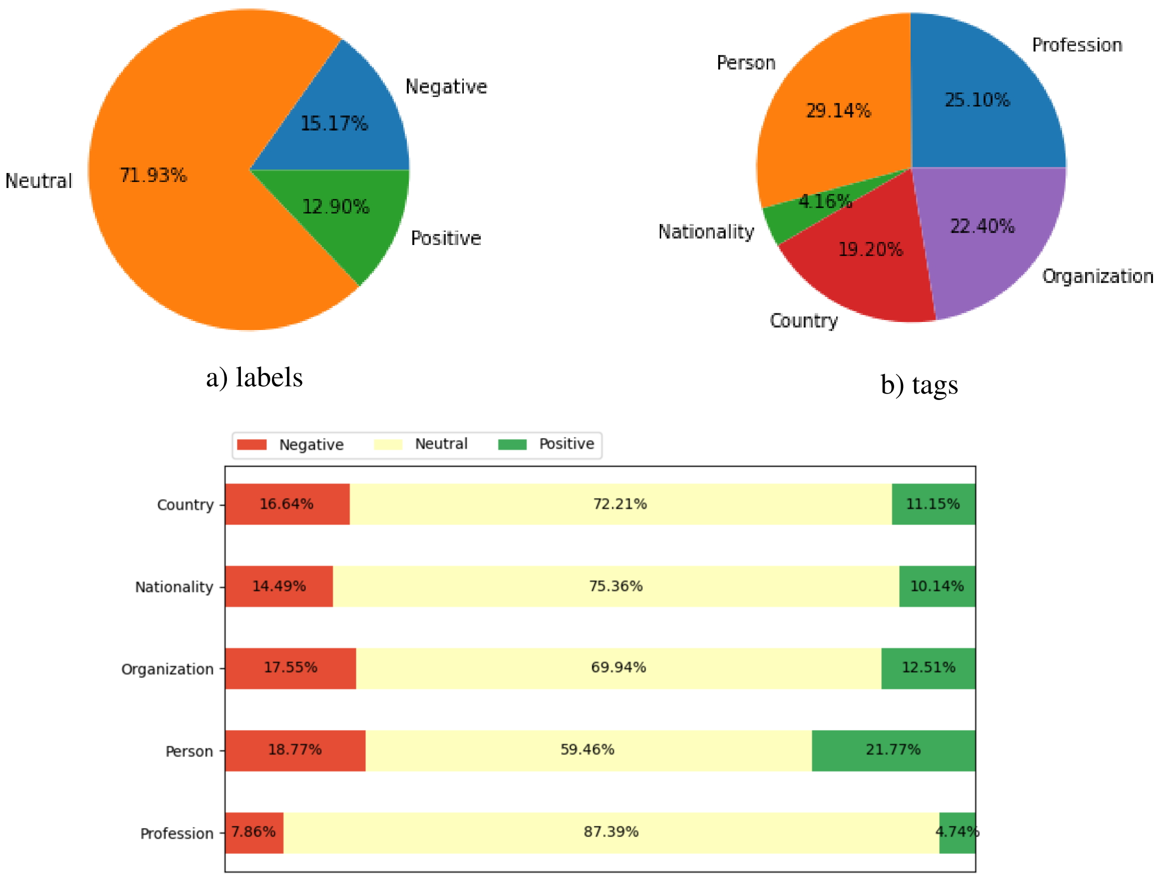
The purpose of the task is to identify sentiments for named entities. The task belongs to the class of targeted sentiment analysis tasks. Based on (Mutlu and Özgür, 2022), the problem of targeted sentiment analysis can be defined as follows. Let $E$ denote all entities in a document $D$. Each $e$ indicates an entity, $E = \{e_1, ..., e_l\}$, $l \in Z^+$. $D = \{w_1, ..., w_k\}$, $k \in Z^+$, where $w$ denotes a word. The objective of targeted sentiment analysis is to find all sentiment pairs $(s_i, t_i)$ in document $D$ where $t_i$ is a target from $T, T = \{t_1, ..., t_m\}$, $t_i \in E$, $m, i \in Z^+$, and $s_i$ is the sentiment toward $t_i$.

The dataset provided for the task contains sentences from mass-media news texts in Russian. Each sentence is annotated by:

- *entity*, the object of sentiment analysis;
- *entity_tag*, the tag for the entity (Country, Nationality, Organization, Person, or Profession);
- *entity_pos_start_rel*, *entity_pos_end_rel*, the indices of the initial and next symbols for the entity occurrence;
- *label*, the sentiment label (negative, neutral, or positive)

Figure 1 shows the distribution of the labels and entity tags in the training set. As can be seen from the figure, most of the entries (71.93%) relate to the neutral class. Some tags are also dominant over others. The most common tag is Person (29.14%), while Nationality is the least abundant (4.16%). The distribution of labels within tags also varies. The texts with the tag Person include the largest proportion of sentiment labels (positive and negative, 40.54%). The smallest proportion of sentiment labels is contained in the tag Profession (12.6%). The breakdown between the training, development, and test sets is shown in Table 1. The number of tokens is obtained using the tokenizer of RuRoBERTa-large[1].

[1] `https://huggingface.co/sberbank-ai/ruRoberta-large`




a) labels

b) tags

c) labels per tags

Figure 1: The distribution of the labels and tags in the training set.

# 4 Methods

## 4.1 Entity Representation

Following previous research (Zhou and Chen, 2022; Alibaeva and Loukachevitch, 2022), we compared several groups of entity representation methods.

- **Entity mask**. This type of entity representation introduces new special tokens for masking the named entity in the source text. We compared two ways to implement this technique. In the first case, we replaced all target entities with a special token *[NE]*. In the second case, we used special tokens *[TYPE]*, where *TYPE* denotes one of the five entity tags.
- **Entity markers**. This representation type introduces new special tokens *[NE]* and *[/NE]* to enclose the named entity. We experimented with the use of one token to enclose the named entity (*[NE] entity [NE]*), as well as two tokens (*[NE] entity [/NE]*).
- **Entity markers (punct)**. This technique encloses the named entity using punctuation (* *entity* *). In this case, we did not introduce new special tokens into the model's vocabulary. The variant of this technique is adding entity types without introducing special tokens (* @ *TYPE* @ *entity* *).
- **Typed entity markers**. This technique is similar to the previous ones, but it uses control codes to highlight named entities. We consider the four types of typed entity markers: replacing target entities with the control code <|*NE*|>; enclosing entities with two similar codes (<|*NE*|>*entity*<|*NE*|>); enclosing entities with different codes (<|*NE*|>*entity*<|*/NE*|>); adding entity types to the control

code (<|*NE:TYPE*|>*entity*<|*NE:TYPE*|>).

All entity representation types are illustrated in Table 2 on the example of the text "Apple и Samsung нарушали патенты друг друга" (*Apple and Samsung infringed on each other's patents*) from the official training set of RuSentNE. In this entry, the target named entity is "Samsung", the entity type is Organization and the sentiment label is -1 (negative).

### 4.2 Models

We compared three pre-trained language models for the Russian language on the named entity oriented sentiment analysis task.

- **RuBERT-base**[2] (Kuratov and Arkhipov, 2019), a BERT-based model for the Russian language with 180M parameters trained on the Russian part of Wikipedia and news data. A multilingual version of BERT-base (Devlin et al., 2019) was used as an initialization.
- **RuBERT-large**[3], a large version of RuBERT containing 427M parameters trained on the Russian part of Wikipedia, news texts, books, and a fragment of the Taiga corpus (Shavrina and Shapovalova, 2017).
- **RuRoBERTa-large**[4], a modification of RuBERT that is pre-trained using dynamic masking (Liu et al., 2019), 355M parameters.

### 4.3 Handling Class Imbalance

Since news texts contain numerous named entities with a neutral sentiment, the neutral class largely dominates in the training set. We experimented with the following methods to reduce the impact of class imbalance on classification performance.

- **Weighted Inverse of Number of Samples (WINS)**, a class weighting technique that weights the samples as the inverse of the class frequency for the class they belong to and then normalizes them over different classes. The weight for the particular class ($w_j$) is calculated as follows:

$$w_j = \frac{n}{c \cdot n_j}, \tag{1}$$

  where $n$ is the number of entries in the dataset, $c$ is the number of classes, $n_j$ is the number of samples of the particular class.
- **Effective Number of Samples (ENS)** (Cui et al., 2019), a class weighting scheme that calculated the weight for a particular class as follows:

$$w_j = \frac{n}{c \cdot E_{n_j}}, E_{n_j} = \frac{1 - \beta^{n_j}}{1 - \beta}, \tag{2}$$

  where $E_{n_j}$ represents the Effective number of Samples, $\beta$ is a hyperparameter ($\beta \in [0, 1)$). We experimented with the $\beta$ values equal to 0.999 and 0.9999.
- **Random Oversampling for Classes (RO$_c$)**, the technique which consists if that randomly selecting entries from minority classes and adding them to the training set until the classes become the same size.
- **Data Augmentation (DA)**, we used back translation (Sennrich et al., 2016) as a data augmentation technique. For each entry from the minority classes we produced new training examples using the public translation engine, Google Translate[5], and the deep-translator Python tool[6].

### 4.4 Resampling Entity Tags

Since the number of entities of different types is not the same, we also investigated resampling methods to balance the number of entity tags. The following approaches were evaluated:

[2] https://huggingface.co/DeepPavlov/rubert-base-cased
[3] https://huggingface.co/sberbank-ai/ruBert-large
[4] See footnote 1
[5] https://translate.google.com/
[6] https://github.com/nidhaloff/deep-translator

<table>
<tr><th>№</th><th>Input representation</th><th>Example</th></tr>
<tr><td>1</td><td>Entity mask - Replacement</td><td>Apple и [NE] нарушали патенты друг друга (Apple and [NE] infringed on each other’s patents)</td></tr>
<tr><td>2</td><td>Entity mask - Type</td><td>Apple и [ORGANIZATION] нарушали патенты друг друга (Apple and [ORGANIZATION] infringed on each other’s patents)</td></tr>
<tr><td>3</td><td>Entity markers - 1</td><td>Apple и [NE] Samsung [NE] нарушали патенты друг друга (Apple and [NE] Samsung [NE] infringed on each other’s patents)</td></tr>
<tr><td>4</td><td>Entity markers - 2</td><td>Apple и [NE] Samsung [/NE] нарушали патенты друг друга (Apple and [NE] Samsung [/NE] infringed on each other’s patents)</td></tr>
<tr><td>5</td><td>Entity markers (punct)</td><td>Apple и * Samsung * нарушали патенты друг друга (Apple and * Samsung * infringed on each other’s patents)</td></tr>
<tr><td>6</td><td>Entity markers (punct) - Type</td><td>Apple и * @ ORGANIZATION @ Samsung * нарушали патенты друг друга (Apple and * @ ORGANIZATION @ Samsung * infringed on each other’s patents)</td></tr>
<tr><td>7</td><td>Typed entity markers - Replacement</td><td>Apple и <|NE|> нарушали патенты друг друга (Apple and <|NE|> infringed on each other’s patents)</td></tr>
<tr><td>8</td><td>Typed entity markers - 1</td><td>Apple и <|NE|>Samsung<|NE|> нарушали патенты друг друга (Apple and <|NE|>Samsung<|NE|> infringed on each other’s patents)</td></tr>
<tr><td>9</td><td>Typed entity markers - 2</td><td>Apple и <|NE|>Samsung<|/NE|> нарушали патенты друг друга (Apple and <|NE|>Samsung<|/NE|> infringed on each other’s patents)</td></tr>
<tr><td>10</td><td>Typed entity markers - Type</td><td>Apple и <|NE:ORGANIZATION|>Samsung <|NE:ORGANIZATION|> нарушали патенты друг друга (Apple and <|NE:ORGANIZATION|>Samsung<|NE:ORGANIZATION|> infringed on each other’s patents)</td></tr>
</table>

Table 2: Types of entity representation.

| № | RuBERT-base | RuBERT-large | RuRoBERTa-large |
|---|---|---|---|
| 1 | 65.29 | 69.64 | 73.17 |
| 2 | 67.34 | 70.65 | 71.28 |
| 3 | 65.25 | **71.66** | 71.88 |
| 4 | 67.25 | 69.95 | 73.18 |
| 5 | 66.1 | 70.72 | **73.62** |
| 6 | 66.52 | 70.33 | 72.71 |
| 7 | 67.1 | 70.42 | 72.56 |
| 8 | **68.15** | 70.12 | 73.3 |
| 9 | 65.99 | 70.37 | 73.16 |
| 10 | 65.99 | 70.32 | 73.27 |

Table 3: Comparison of entity representations and models (macro F1-score, %).

- **Random Oversampling for Tags (RO$_t$)**, the technique is similar to Random Oversampling for Classes, but the purpose is to balance the number of entries with different tags.
- **Sentence-Level Resampling (SLR)** (Wang and Wang, 2022), the technique was proposed for named entity recognition to increase the number of tokens of a particular entity type in the training set. The resampling function $f_s$ can be adapted for our task in the following way. Let us denote the set of all target entity tags as $T$. Let $c(t, s)$ be the number of tokens of the target named entity in sentence $s$. The rareness $r$ of the entity tag is measured as follows:

$$r_t = -log_2 \frac{\sum_{s \in S} c(t, s)}{N}, \tag{3}$$

where $S$ is the set of all sentences in the training set, $\sum_{s \in S} c(t, s)$ is the total number of tokens included in the target named entities with the type $t$ in the training set, $N$ is the number of all tokens in the training set.

$$f_s = \frac{r_t \cdot \sqrt{c(t, s)}}{\sqrt{l_s}}, \tag{4}$$

where $l_s$ is the number of tokens in the particular text. The resampling function $f_s$ shows the number of times a sentence $s$ should be resampled in a training set. The greater the number of tokens of the target entity, and the less often the entity tag is presented in the training set, the more frequently the sentence is resampled.

## 5 Results

### 5.1 Development Phase

During the development phase, we evaluated the techniques presented in Section 4. The training set was split into training and validation subsets in a ratio of 70:30. We fine-tuned each model for 6 epochs with a learning rate of 5e-6, a maximum sequence length of 200, and a batch size of 8. To evaluate the results on the validation subset, we used the macro F1-score.

Table 3 presents the results of the comparison of the models and entity representations. The highest scores for each model are shown in bold. The three best results across all models are highlighted. For better presentation, a correspondence between the types of entity representation utilized in this work and their sequential numbers is listed in Table 2. RuRoBERTa-large demonstrated the highest scores across all entity representation types. None of the entity representations showed a clear advantage over others. For instance, entity representation type 3 (Entity markers - 1) demonstrated the highest F1-score for RuBERT-large (71.66%) and the lowest for RuBERT-base (65.25%). For RuBERT-base, the best result was obtained using entity representation type 8 (Typed entity markers - 1). For RuRoBERTa-large,

| Technique | RuRoBERTa-large (5) | RuRoBERTa-large (8) | RuRoBERTa-large (10) |
|---|---|---|---|
| WINS | **73.83** ↑ | **73.44** ↑ | 72.8 |
| $\text{ENS}_{\beta=0.999}$ | 73.18 | **74.42** ↑ | **73.62** ↑ |
| $\text{ENS}_{\beta=0.9999}$ | 72.64 | 72.49 | 71.84 |
| $\text{RO}_c$ | 73.32 | 71.75 | 72.37 |
| DA | 72.46 | 72.98 | 71.71 |
| $\text{RO}_t$ | 73.51 | 72.8 | 72.81 |
| SLR | **74.23** ↑ | **73.34** ↑ | 71.94 |

Table 4: Comparison of strategies for handling class imbalance and resampling entity tags (macro F1-score, %).

the highest score was achieved with entity representation type 5 (Entity markers (punct)). Since many models showed very similar results, we selected three models with the highest values of the F1-score for further experiments. The selected models include Ru-RoBERTa-large with entity representation types 5 (Entity markers (punct), 73.62% of F1-score), 8 (Typed entity markers - 1, 73.3%), and 10 (Typed entity markers - Type, 73.27%).

In Table 4, the results for comparing strategies for class weighting and resampling entity tags are presented. The numbers of the corresponding entity representation types are given in brackets in the names of the columns. The results that exceeded the result of the corresponding model without the use of the strategy are shown in bold and marked with an arrow (↑). It can be seen from the values in the table that no strategy gave an advantage on all compared models. WINS showed a slight improvement with the entity representations 5 (+0.21%) and 8 (+0.14%). ENS with the value of $\beta$ equal to 0.9999 ($\text{ENS}_{\beta=0.9999}$) increased the RuRoBERTa-large performance using the entity representation types 8 (+1.12%) and 10 (+0.35%). Other strategies for handling class imbalance ($\text{ENS}_{\beta=0.999}$, $\text{RO}_c$, and DA) led to a performance decrease in our experiments. Concerning the issue of resampling entity tags, $\text{RO}_t$ worsened scores for all the considered models while SLR increased the F1-score with the entity representation types 5 (+0.61%) and 8 (+0.04%).

The development phase showed that the results may vary depending on the type of entity representation. Nevertheless, in our experiments, the best result for each entity representation type was achieved by RuRoBERTa-large. The choice of the entity representation type may not be obvious due to the close results obtained by models. Some strategies for handling class imbalance and resampling entity tags demonstrated an improvement in performance on the validation subset. However, not a single strategy showed an increase for all models and the growth value was often small. Therefore, during the phase, several strategies and entity representation types were selected for use in the evaluation phase.

### 5.2 Evaluation Phase

During this phase, we experimented with the models fine-tuned, using the techniques that showed an improvement in the development phase (WINS, $\text{ENS}_{\beta=0.9999}$, SLR) and the entity representations that demonstrated the best results during the development phase (5, 8, and 10). To increase the results on the test set, we also utilized ensemble learning and produced silver labels for the unlabelled development set provided by the organizers. Our best submission for the evaluation phase represents a system based on RuRoBERTa-large, fine-tuned using WINS with entity representation type 8. We utilized the augmented dataset consisting of the official training set and the development set with silver labels. The total size of the augmented dataset was 9,482. To combine the predictions of fine-tuned models, we used a soft-voting technique.

The official results are presented in Table 5. The models were evaluated in terms of the macro $\text{F1}_{p,n}$-score (the main performance metric), which is averaged over two sentiment classes, and the macro $\text{F1}_{p,n,0}$-score for three-class classification. Our system demonstrated the best $\text{F1}_{p,n,0}$-score out of nine

| Score | Metric | |
|---|---|---|
| | $F1_{p,n}$-score | $F1_{p,n,0}$-score |
| F1-score, % | 66.64 | 74.29 |
| rank | 2 | 1 |
| baseline | 40.92 | 56.71 |
| avg F1-score | 58.27 | 67.12 |

Table 5: Official results.

submitted teams and the second $F1_{p,n}$-score (0.03% from the first-place team).

## 6 Error Analysis

In this section, we provide some error examples produced by RuRoBERTa-large, fine-tuned on the official development set. The gold labels for the development set were released by the organizers of the evaluation after the end of RuSentNE. Since the gold labels for the test set had not been published by the time this paper was submitted, we cannot analyse the errors of our final model. However, an analysis of errors on the development set makes it possible to empirically trace the general trends.

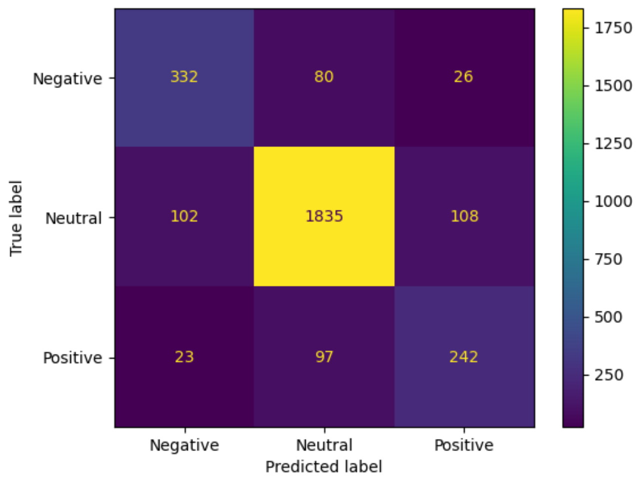



Figure 2: Confusion matrix (the development set).

The confusion matrix for the development set is presented in Figure 2. As can be seen from the figure, most of all errors are associated with the classifying entries from the neutral class as positive or negative. Examples of such errors are given in Table 6 (1 and 2). In the first sentence, the model predicts a positive class, probably, due to the availability of positive information ("universally recognized record"), however, the general meaning of the sentence is interpreted incorrectly. Perhaps this error is because the sentence is quite long and contains co-reference expressions ("Jeanne Calment", "who", "whose"). The second example is classified as negative in view of the presence of a negative fact ("was deprived of all victories"). Sentences 3-5 in Table 6 illustrate the opposite situation when sentences from the sentiment classes are classified as neutral. In the third sentence, as in the first, the presence of co-reference expressions ("Rusnok", "new prime minister") leads to an error. In addition, in examples 3 and 4, the lack of knowledge of the context complicates the classification. Examples 2 and 5 look thematically similar, but they contain entities with different tags (Person and Profession respectively). Finally, sentences 6 and 7 illustrate the situation when the model predicts the opposite sentiment class.

| № | Sentence | Predicted label | Actual label |
|---|---|---|---|
| 1 | Общепризнанный рекорд долголетия принадлежит француженке Жанне Кальман, скончавшейся в 1997 году в возрасте 122 лет и 164 дней, возраст которой подвергается сомнению (*The universally recognized record of longevity belongs to a French woman Jeanne Calment, who died in 1997 at the age of 122 and 164 days, whose age is questioned*) | Positive | Neutral |
| 2 | Лэнса Армстронга лишили всех побед на "Тур де Франс" (Lance Armstrong was deprived of all victories in the "Tour de France") | Negative | Neutral |
| 3 | Земан назначил Руснока под предлогом, что новый премьер - хороший экономист, который займется подготовкой бюджета следующего года (*Zeman appointed Rusnok under the pretext that the new prime minister is a good economist who will engage in the preparation of the budget for the next year*) | Neutral | Positive |
| 4 | Через Германию пролегали маршруты нелегальных самолётов, которые перевозили заключённых (*The routes of illegal aircraft that transported prisoners ran through Germany*) | Neutral | Negative |
| 5 | Восемь бадминтонисток были дисквалифицированы на Олимпийских играх (*Eight badmintonists were disqualified at the Olympic Games*) | Neutral | Negative |
| 6 | Россия и Китай заблокировали резолюцию ООН, направленную против правительства Сирии (*Russia and China blocked the UN resolution directed against the Government of Syria*) | Negative | Positive |
| 7 | Лебедев признал свое участие в драке, но отверг обвинения в хулиганстве и политической ненависти. (*Lebedev admitted his participation in a fight, but rejected accusations of hooliganism and political hatred*) | Positive | Negative |

Table 6: Error examples (the development set). The target entity is highlighted.

In general, in such cases, the model pays more attention to the nearest context of the entity, without analysing the general meaning of the sentence.

## 7 Conclusion

In this paper, we present our approach to performing named entity oriented sentiment analysis of Russian news texts. The proposed method is based on the use of RuRoBERTa-large using class weighting, data augmentation with silver data, and ensemble learning. We also studied the impact of the use of different entity representation types and strategies for handling class imbalance and resampling the dataset and provided the results of error analysis. We foresee two directions for future work. One potential direction is to investigate the impact of co-reference resolution as a pre-processing step for named entity sentiment analysis of Russian texts. Another future direction is exploring approaches for the inclusion of contextual-semantic information.

## References


Kamila Alibaeva and Natalia Loukachevitch. 2022. Analyzing COVID-related Stance and Arguments using BERT-based Natural Language Inference. // *Computational Linguistics and Intellectual Technologies*, P 8–16.

Pavel Blinov and Evgeniy V Kotelnikov. 2015. Semantic similarity for aspect-based sentiment analysis. *Russ. Digit. Libr. J.*, 18(3-4):120–137.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. 2019. Class-balanced loss based on effective number of samples. // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, P 9268–9277.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, P 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Athanasios Giannakopoulos, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl. 2017. Unsupervised Aspect Term Extraction with B-LSTM & CRF using Automatically Labelled Datasets. *EMNLP 2017*, P 180.

Anton Golubev, Nicolay Rusnachenko, and Natalia Loukachevitch. 2023. RuSentNE-2023: Evaluating entity-oriented sentiment analysis on Russian news texts. // *Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue"*.

V.V. Ivanov, E.V. Tutubalina, N.R. Mingazov, and I.S. Alimova. 2015. Extracting aspects, sentiment and categories of aspects in user reviews about restaurants and cars. // *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, P 22–33.

Evgeny Kotelnikov, Natalia Loukachevitch, Irina Nikishina, and Alexander Panchenko. 2022. RuArg-2022: Argument Mining Evaluation. // *Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue"*.

Anastasia Kotelnikova, Danil Paschenko, Klavdiya Bochenina, and Evgeny Kotelnikov. 2022. Lexicon-based methods vs. BERT for text sentiment analysis. // *Analysis of Images, Social Networks and Texts: 10th International Conference, AIST 2021, Tbilisi, Georgia, December 16–18, 2021, Revised Selected Papers*, P 71–83. Springer.

Dilek Küçük and Fazli Can. 2021. Stance detection: Concepts, approaches, resources, and outstanding issues. // *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, P 2673–2676.

Yuri Kuratov and Mikhail Arkhipov. 2019. Adaptation of deep bidirectional multilingual transformers for Russian language. // *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, P 333–339.

Hao Li and Wei Lu. 2017. Learning latent sentiment scopes for entity-level sentiment analysis. // *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Fei Liu, Trevor Cohn, and Timothy Baldwin. 2018. Recurrent Entity Networks with Delayed Memory Update for Targeted Aspect-based Sentiment Analysis. // *Proceedings of NAACL-HLT*, P 278–283.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Natalia Loukachevitch, Pavel Blinov, Evgeny Kotelnikov, Yulia Rubtsova, Vladimir Ivanov, and Elena Tutubalina. 2015. SentiRuEval: testing object-oriented sentiment analysis systems in russian. // *Proceedings of International Conference Dialog*, volume 2, P 3–13.

Natalia Loukachevitch. 2021. Automatic sentiment analysis of texts: the case of Russian. *The Palgrave Handbook of Digital Russia Studies*, P 501–516.

Nikita Lozhnikov, Leon Derczynski, and Manuel Mazzara. 2020. Stance prediction for Russian: data and analysis. // *Proceedings of 6th International Conference in Software Engineering for Defence Applications: SEDA 2018 6*, P 176–186. Springer.

Yukun Ma, Haiyun Peng, and Erik Cambria. 2018a. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. // *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Yukun Ma, Haiyun Peng, Tahir Khan, Erik Cambria, and Amir Hussain. 2018b. Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis. *Cognitive Computation*, 10:639–650.

Iuliia Makogon and Igor Samokhin. 2022. Targeted sentiment analysis for Ukrainian and Russian news articles. // *ICTERI 2021 Workshops: ITER, MROL, RMSEBT, TheRMIT, UNLP 2021, Kherson, Ukraine, September 28–October 2, 2021, Proceedings*, P 538–549. Springer.

V. Mayorov, I. Andrianov, N. Astrakhantsev, V. Avanesov, I. Kozlov, and D. Turdakov. 2015. A high precision method for aspect extraction in Russian. // *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, P 34–43.

Martin Müller, Marcel Salathé, and Per E Kummervold. 2020. COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter. *arXiv preprint arXiv:2005.07503*.

Mustafa Melih Mutlu and Arzucan Özgür. 2022. A Dataset and BERT-based Models for Targeted Sentiment Analysis on Turkish Texts. // *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, P 467–472.

Aleksandr Naumov, R Rybka, A Sboev, A Selivanov, and A Gryaznov. 2020. Neural-network method for determining text author's sentiment to an aspect specified by the named entity. // *CEUR Workshop Proceedings*, P 134–143.

Eduard Nugamanov, Natalia Loukachevitch, and Boris Dobrov. 2021. Extracting sentiments towards COVID-19 aspects. // *Supplementary 23rd International Conference on Data Analytics and Management in Data Intensive Domains, DAMDID/RCDL 2021*, P 299–312.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. // *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, P 2227–2237, New Orleans, Louisiana, June. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammed AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. // *ProWorkshop on Semantic Evaluation (SemEval-2016)*, P 19–30. Association for Computational Linguistics.

Y.V. Rubtsova and S.A. Koshelnikov. 2015. Aspect Extraction Using Conditional Random Fields. *SentiRuEval-2015*.

Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. 2016. SentiHood: Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods. // *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, P 1546–1556.

Svitlana Salnikova and Roman Kyrychenko. 2021. Sentiment analysis based on the BERT model: Attitudes towards politicians using media data. // *Proceedings of the International Conference on Social Science, Psychology and Legal Regulation (SPL 2021)*, P 39–44. Atlantis Press.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Edinburgh Neural Machine Translation Systems for WMT 16. // *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, P 371–376.

Esther Irawati Setiawan, Ferry Ferry, Joan Santoso, Surya Sumpeno, Kimiya Fujisawa, and Mauridhi Hery Purnomo. 2020. Bidirectional GRU for targeted aspect-based sentiment analysis based on character-enhanced token-embedding and multi-level attention. *Computing*, 1:2.

Tatiana Shavrina and Olga Shapovalova. 2017. To the methodology of corpus construction for machine learning: "Taiga" syntax tree corpus and parser. // *Proceedings of "CORPORA-2017" International Conference*, P 78–84.

Sergey Smetanin. 2020. The applications of sentiment analysis for Russian language texts: Current challenges and future perspectives. *IEEE Access*, 8:110693–110719.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. // *Proceedings of NAACL-HLT*, P 380–385.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

Denis Tarasov. 2015. Deep recurrent neural networks for multiple language aspect-based sentiment analysis of user reviews. // *Proceedings of the 21st international conference on computational linguistics dialog*, volume 2, P 53–64.

V.G. Vasilyev, A.A. Denisenko, and D.A. Solovyev. 2015. Aspect extraction and twitter sentiment classification by fragment rules. // *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, P 76–86.

Sergey V. Vychegzhanin and Evgeny V. Kotelnikov. 2017. Stance Detection in Russian: a Feature Selection and Machine Learning Based Approach. // *AIST (Supplement)*, P 166–177.

Sergey V. Vychegzhanin and Evgeny V. Kotelnikov. 2019. Stance detection based on ensembles of classifiers. *Programming and Computer Software*, 45:228–240.

Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z Pan. 2020. Target-aspect-sentiment joint detection for aspect-based sentiment analysis. // *Proceedings of the AAAI conference on artificial intelligence*, volume 34, P 9122–9129.

Xiaochen Wang and Yue Wang. 2022. Sentence-level resampling for named entity recognition. // *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, P 2151–2165.

Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, and Yongfeng Huang. 2018. A hybrid unsupervised method for aspect term and opinion target extraction. *Knowledge-Based Systems*, 148:66–73.

Hu Xu, Bing Liu, Lei Shu, and S Yu Philip. 2018. Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction. // *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, P 592–598.

Yunyi Yang, Kun Li, Xiaojun Quan, Weizhou Shen, and Qinliang Su. 2020. Constituency lattice encoding for aspect term extraction. // *Proceedings of the 28th international conference on computational linguistics*, P 844–855.

Zhihao Ye and Zhiyong Li. 2020. A variant of recurrent entity networks for targeted aspect-based sentiment analysis. // *ECAI 2020*, P 2268–2274. IOS Press.

Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. // *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, P 2979–2985.

Wenxuan Zhou and Muhao Chen. 2022. An improved baseline for sentence-level relation extraction. // *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, P 161–168.