Adaptation of Semantic Rule-Based Sentiment Analysis Approach for Russian Language

Ilya Paramonov, Anatoliy Poletaev
P.G. Demidov Yaroslavl State University
Yaroslavl, Russia
ilya.paramonov@fruct.org, anatoliy-poletaev@mail.ru

***Abstract*—The paper describes application of the semantic rule-based sentiment analysis approach, which was earlier developed and tested on English texts, to the Russian language. In order to take into account specificity of Russian it was adapted, particularly representation of the rules as patterns over a list of words was replaced with algorithms over the syntax tree of a sentence. The experiments on a quarter of a corpus of sentences extracted from hotel reviews allowed to perform the error analysis and refinement of the approach. The final results on the whole corpus allowed to achieve the results close to the state-of-the-art methods based on neural networks. The advantages of the approach, including simple interpretability of its results and absence of the need of learning, make it perspective for further research in sentiment analysis.**

I. Introduction

This paper is dedicated to sentiment analysis of Russian sentences. We consider the task of sentence-level sentiment analysis as identifying the sentence author's attitude to the topic of the sentence. A sentence is *positive* if it contains a positive fact, opinion, or emotion expressed by the author and there are no negative facts, opinions, or emotions or these negative expressions are overlapped by positive ones. In the opposite situation (when negative facts, opinions, or emotions prevail) the sentence is *negative*. If the sentence is neither positive nor negative we consider it to be *neutral* [1], [2].

Most of the cutting-edge sentiment analysis approaches both for Russian and English utilize neural networks, such as BERT and LSTM [3]–[5]. Neural networks are powerful tools that allow researchers to practically process texts in a natural language with minimum assumptions on how it works.

Although less common than neural networks, there are approaches based on semantic or syntactic rules aimed at handling the ways of how sentiments are expressed in a sentence [6]–[10]. Despite the fact that creating such rules requires detailed language study, the rule-based approaches in some cases can surpass neural networks drawbacks: semantic rules do not need large corpora to be trained on and their results are rather easy to understand, which makes error analysis easier.

Nowadays, semantic rules-based approaches are underdeveloped and even somewhat neglected for Russian language. For example, the authors of one of the most recent comprehensive surveys on sentiment analysis in Russian [5] just state that machine learning based approaches outperform rule-based approaches but do not provide any comparison of their performance. In this paper we aim to fill this gap.

The main goal of this work is to explore the possibility of reaching the state-of-the-art results using a rule-based sentiment analysis approach for sentences in Russian. We base our work on the approach described by O. Appel et al. [9]. We adapt it for Russian language and evaluate its performance on a hotel reviews corpus. The original approach utilizes semantic rules implemented as patterns applied to a list of words representing a sentence. However, it is difficult to apply semantic patterns in Russian, because unlike English it has no strict word order. To resolve this issue, we reconstruct the rules as algorithms over the syntax tree of a sentence.

The rest of this paper is structured as follows. Section II describes the related work. In Section III we give a general description of the proposed semantic rules-based approach in comparison with the baseline work. Section IV is devoted to a detailed description of the adapted semantic rules for Russian language. Section V contains the results of experiments, the description of the approach refinement, and comparison with the state-of-the-art results. In conclusion we summarize the results and propose ideas for future works.

II. Related work

The usage of semantic rules for sentence-level sentiment analysis was originally proposed by M. Shaikh et al. [11] and improved with usage of types of words dependencies by L. Tan et al. [2]. These rules were aimed at determining the sentiment expressed in a single clause.

Y. Xie et al. [8] introduced an advanced rule-based approach that also takes different kinds of clauses into consideration. The approach is based on the assumption that the sentiment of a sentence is determined by the majority sentiment of all the sentence words. Xie introduced 13 semantic rules aimed at handling different ways of how sentiment can be expressed in a sentence. There were two methods of sentiment analysis proposed: by the prevailing sentiment and by using a machine learning method (decision tree, neural network, logistic regression, or random forest).

The performance of the approach for English was evaluated on two datasets: 1000 Facebook comments and 500 tweets. All the sentences from these datasets were manually annotated as positive, negative, or neutral. The following results were obtained:

- Facebook comments dataset: by prevailing sentiment F_1 = 0.76; using the random forest, which gave the best performance among all the used machine learning algorithms, F_1 = 0.92.
- Twitter tweets dataset: by prevailing sentiment F_1 = 0.68; using the neural network, which gave the best performance among all the used machine learning algorithms, F_1 = 0.81.

This sentiment analysis approach was improved by O. Appel et al. [9]. Their implementation uses 10 semantic rules, 8 of them were taken from the original Xie's work and 2 were introduced by the authors of the research. The authors also proposed semi-automatic sentiment dictionary enrichment using machine learning. The idea of this technique is automatic listing all the words, which can have sentiment but are not present in the dictionary. For example, all the words from the positive sentence falsely determined as negative or neutral become candidates for inclusion to the dictionary as positive words. Finally, the list of candidates is manually processed by the human expert who finally adds each candidate to the dictionary or reject it. The sentiment analysis method is used twice on the same corpus, the second time with the dictionary enriched.

The implemented approach was evaluated by its authors on three datasets. On the first and the second ones consisting of Twitter comments it achieves the accuracy of approximately 0.88. On the third dataset consisting of movie reviews the accuracy is about 0.76.

These experiments show that the semantic rule-based approach can achieve rather high sentiment analysis quality. The main drawback of all the aforementioned works is the lack of error analysis that lefts unclear a number of important questions, e.g., whether the sentiment analysis quality can be improved by introducing new semantic rules and how often they are applied.

Construction of semantic rules is a rather complicated task that requires deep understanding of the language structure, that is why nowadays this class of methods is mostly supplanted by neural network-based methods. Modern neural network architectures, such as BERT, LSTM, and GCN are able to determine the sentiments of sentences with rather high precision. According to [3]–[5] F_1-measure of 0.75–0.85 can be achieved with the BERT architecture that is state-of-the-art.

However, the fact that a neural network is in fact a black box leads to the lack of interpretation possibilities of the results. As a consequence, it is impossible to reason why the neural network performance changes significally between different corpora. For example, RuBERT neural network shows F_1 = 0.64 for the SentiRuEval-2015 TC corpus and F_1 = 0.77 for the RuReviews corpus [5]. Another pitfall of the neural networks usage is a necessity of huge annotated corpora to train the neural network.

All these works are devoted to the sentiment analysis of sentences in English. The most recent work involving such an approach for Russian language was published in 2013 [12] when such approaches were not well-developed. It led to rather low quality of sentiment analysis with F_1-measure of approximately 0.50.

In a nutshell, the quality measures for syntactic rules based and neural networks based approaches can be close. For example, on the same movie reviews dataset the aforementioned semantic rule-based approach [9] achieved F_1-measure of 0.76, whereas the result of BERT [13] was 0.79. That is why it looks reasonable to revive the interest to this approach especially for Russian language, for which it was underused, taking into account its advantages of simple interpretability of results and error analysis.

III. Rule-based sentiment analysis approach and its adaptation to the Russian language

A. General scheme

In the proposed approach the word is considered as having a sentiment (i.e., be a sentiment bearer) or having no sentiment. To identify the sentiment, sentiment scores are used. The sentiment scores are three decimal numbers between zero and one, sum of which is equal to one. The first and the second ones show how positive and negative the word is. The third represents the neutral component of the word sense, i.e., to what extent the word expresses facts independent from the author's attitude.

Following [9], we process a sentence to determine its sentiment according to the following steps:

1) Initialization. The initial sentiment scores are taken from the sentiment dictionary. All words that are not present in the dictionary are considered as neutral.
2) Semantic rules application. For each word the rules are applied in a specific order, each of them can modify the sentiment scores. The details are provided in Section IV.
3) Final determination of the words sentiments. Each word is considered to be positive if the positive sentiment score is the highest, negative if the negative sentiment score is the highest, and neutral otherwise.

After all these steps, the sentiment of a sentence is determined by the majority sentiment of all the sentence words. If there are more positive words in the sentence than negative ones, the sentence is considered as positive; if there are more negative than positive words in the sentence, it is determined as negative; otherwise, it is considered neutral.

B. Adaptation to the Russian language

In [9] sentences are processed as lists of words. Each word in the list has its part-of-speech tag. The rules are implemented as patterns: if a part of the list matches a pattern, the corresponding rule is applied. Although the approach is declared as multilingual, it relies on the word order in a sentence too much to be used for Russian, which has no strict word order.

For example, consider the sentence *Нам принесли чистые полотенца* (*We were provided with clean towels*). We can change the word order without changing its sense. For example, the word '*полотенца*' can be placed at almost every place in the sentence, e.g., *Полотенца нам принесли чистые*.

Hence, it is hard to propose a syntactic pattern to process all the combinations with the positive word '*чистые*'.

This issue has already been noticed in the work on adaptation of the PULS event extraction framework [14]. Its authors proposed a patterns extension to cover all possible cases of words arrangement. However, such an extension can make patterns extremely complex and it is difficult to follow this approach in practice.

To overcome this issue we refused from processing sentences as lists in favor of using syntax trees, which reflect the syntactic structure of the sentence and not only the word order.

The syntax tree of a sentence is a rooted tree, where each node represents a single word. The sentence's predicate is the root of the syntax tree. Direct children of the word are its dependants, i.e., if c is a dependant of w, then w is the parent of c. The path between the word and the sentence root consists of words ancestors. If w is an ancestor of c, then c is called a descendant of w.

Each edge represents a dependency between the head word and its dependant. Each word has a part-of-speech tag. Each dependency has a type, e.g., the *nsubj* dependency type describes a relation between the predicate head and a subject dependant expressed by a noun. In this paper we use the Universal Dependencies system of part-of-speech tags and dependencies types [15], the latest version is available at https://universaldependencies.org.

In our research we implemented the rules for Russian language as algorithms over the syntax tree. They are described in the following section.

IV. Semantic rules used in the Russian adaptation of the rule-based sentiment analysis approach

This section describes adaptation of the rules from [9] to the Russian language performed in this research. Notation of the rules follows the aforementioned paper.

R1 and R10. Negation handling rules

These rules are aimed at handling negations expressed by a negative particle (R1) or by a comparative clause (R10).

In Russian, the '*не*' particle can negate all the phrase sense and, hence, the sentiment, if it stands before the sentence predicate and negate the sense of the following word otherwise [16]. Distinction of these two cases is a difficult task because of the fact that the predicate in Russian can be expressed by several words of various parts of speech in various grammatical forms. At the moment, the quality of existing syntactic parsers does not allow us to solve this problem. Therefore, we assume that the negative particle negates the sentiments of all the words of a phrase.

To handle the fact that the sense of a phrase in a sentence is negated, we introduced the negation operation. This operation is applied to the sentiment scores of the word and defined as follows:

$$\mathrm{Neg}(pos, neg, obj) = (neg, pos, obj),$$

i.e., it swaps the positive and negative scores of the word leaving its neutral component intact.

The negation operation is applied to all the words in a particular subtree if one of its root dependants is the negative particle.

Fig. 1 shows an example of R1 applying to the sentence *Мы будем жаловаться на то, что нам* ***не*** *положили чистые полотенца* (*We will complain that we were not provided with clean towels*). The '*не*' node (colored heavy grey in the syntax tree) is a child of the '*положили*' node. Hence, we negate sentiments of '*положили*' and all its descendants (colored light grey in the syntax tree). '*Не*' has no sentiment, it only modifies the sentiment of its parent and its descendants.

Negation can be applied twice or more times to handle, for example, double negation, which does not change the phrase sense in Russian.

English comparative clause *not as ... as ...* has exact counterparts in Russian, for example, '*не так ..., как ...*' and '*не ... настолько, насколько ...*'. The pattern of that comparative clause consists of:

- negative particle '*не*',
- adjective of positive degree syntactically connected with one of the comparative words '*так*', '*такой*', '*настолько*'
- noun phrase syntactically connected with one of the comparative conjunctions '*как*', '*какой*', '*насколько*'.

Hence, in our adapted rule we check whether adjectives and nouns are a part of comparative clause described before. If an adjective is a part of such comparative clause, we negate its sentiment unless it was previously negated when applying the R1 rule. If the noun is a part of the a clause, we consider it having no sentiment.

For example, in the sentence *Наш номер был* ***не таким*** *красивым,* ***как*** *соседский* (*Our room was not as beautiful as the our neighbour's one*) the sentiment scores of '*красивым*' are negated and the sentiment of '*соседский*' is not taken into account when calculating the sentiment of the sentence. In Fig. 2 the words, which are shaping the comparative clause framework, are colored heavy grey and the word, which sentiment is negated, is colored light grey.

Adaptation of the R10 rule is a good example of complexity growth when patterns and ideas developed for the strict word order languade are transfered to a language with no strict word order.

R3, R6 and R7. Composition rules

These rules are aimed at handling the fact that some of the word pairs in the sentences are stronly interconnected in their meanings and should be composed.

These word pairs include:

- an action and its actor (R3),
- an action and its adverbal modifier (R6),
- a noun and a participle modifying the noun (R7).

These rules provide another example of a challenge related to the word order. In Russian, unlike English, an action verb

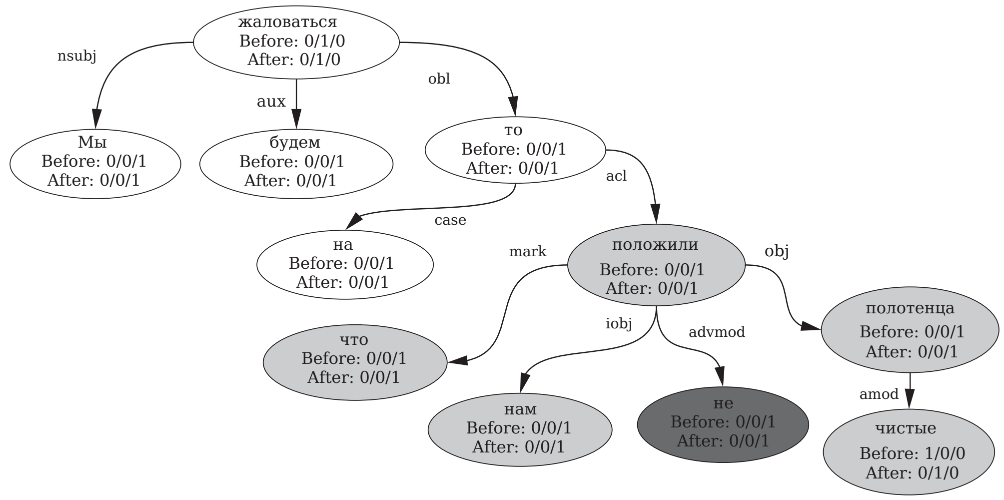

Fig. 1. Sentiment scores of the sentence's words before and after application of the R1 rule

can precede or follow its actor noun. For example, in the sentence *Вентиляционная* ***система не работала*** (*The ventilation system was out of order*) the action verb '*работала*' follows its actor noun '*система*' and in the sentence *В нашем номере* ***не работала система*** *вентиляции (In our room the ventilation system was out of order)* the action verb preceeds its actor noun. Moreover, there can be a constituent or a clause between the action verb and its actor noun. In the sentence *Вентиляционная* ***система*** *в нашем номере* ***не работала*** (*The ventilation system in the room was out of order*) the action verb and the actor noun are separated by '*в нашем номере*' and in the sentence *В нашем номере* ***не работала****, хоть её и пытались починить, вентиляционная* ***система*** *(In our room the ventilation system was out of order despite the fact that the personnel tried to fix it)* there is a dependent clause between the action verb and its actor noun. Similar examples can be found for the other composed pairs.

The composition operation is a calculation of joint sentiment scores of the words with subsequent consideration of these words as a one for the purposes of sentence sentiment detection. It is implementied by assuming one of the composed words (the word that is closer to the syntax tree root) as having no sentiment and assuming the other word sentiment as a joint sentiment of composed words.

If the sentiment scores of composed words are equal, the joint sentiment scores are also equal to them. Otherwise, the sentiment scores of the word with the strongest (the closest to one by the absolute value) word are assumed as the joint sentiment scores. If the words sentiments are equally strong, the averages of their sentiment scores are assumed as the joint sentiment scores.

In the R3 adaptation we search for a parent verb and a dependant noun with a dependency of the *nsubj* type between them and compose the sentiments of the verb and the noun. Similary, in the R6 adaptation we search for a parent noun and a dependant child where '*что*' is one of the verb dependants and in the R7 adaptation we search for a head noun and a dependant with the *VERB* part-of-speech tag with the *amod* dependency between them (Russian participles can be identified by a combination of the *VERB* part-of-speech tag and the *amod* dependency type).

R11–R15. Clauses processing

These rules are aimed at processing various different types of clauses:

- contradiction clauses introduced by 'but' (R11) or 'however' (R15),
- a concessive clause (R12),
- a hypotetical clause (R13),
- a time clause (R14).

All these rules disregard the sentiments of the words in a part of a sentence. R11 and R15 disregard the words outside the processed clause and the R12–R14 rules disregard the words inside the processed clause. If a word is disregarded we assume it as having no sentiment. Whether a word is inside or outside the processed clause and whether a word is affected by the rule is determined by searching over the syntax tree, where a clause corresponds to a subtree.

In the R11 and R15 adaptations we search for a subtree, which root has a child '*но*', '*зато*', or '*однако*'. For example, in the sentence *Отель хороший,* ***но*** *понятливость и учтивость персонала оставляют желать лучшего* (*The hotel is nice, but personnel understanding and courtesy leave much to be desired*) the sentiment scores of the words '*отель*' and '*хороший*' are not taken into account when calculating the sentiment.

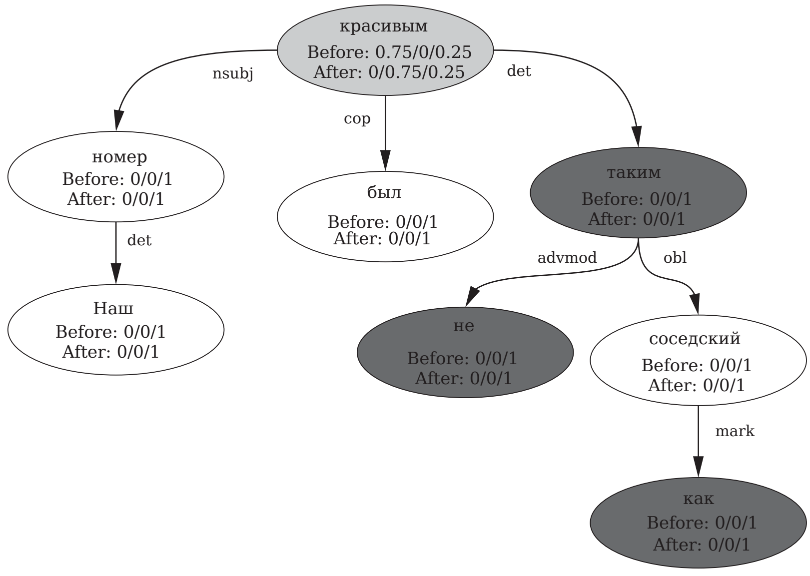

Fig. 2. Sentiment scores of the sentence's words before and after application of the R10 rule

In the R12 adaptation we search for a subtree, which root has a child '*хотя*'. For example, in the sentence *Мне понравился фильм,* ***хотя*** *я и не люблю снявшего его режиссёра* (*I liked the film although I dislike its director*) only the sentiments of the words '*мне*', '*понравился*' and '*фильм*' are used to calculate the sentiment of the sentence.

In the R13 adaptation we search for a subtree that root has a child '*если*' and '*не*'. For example, in the sentence *В следующем году мы обязательно вернёмся в этот прекрасный отель,* ***если не*** *закроют границы из-за пандемии* (*Next year we will definitely return to this nice hotel unless the borders are closed due to the pandemia*) the sentiment scores of words in the clause '*закроют границы из-за пандемии*' are not taken into account when calculating the sentiment of the sentence.

In the R14 adaptation we search for a subtree, which root has a '*пока*' child. For example, in the sentence *Мы прекрасно провели время в лобби-баре,* ***пока*** *ужасный ливень не закончился, и нам так понравилось, что мы вернулись туда вечером* (*We spent great time in the lobby bar, waiting for the terrible downpour to end, and we liked the bar so much that we returned at the evening*) the sentiments of the words '*ужасный*', '*ливень*' and '*не закончился*' are not used for calculating the sentiment of the sentence.

Rules application order

The negation handling rules have to be applied after the composing rules because it is necessary to negate the composed sentiments. The clauses processing rules do not require results of the other rules application and do not provide results to be used by the other rules. Taking these considerations into account we applied the rules in the following order: R3, R6, R7, R1, R10, R11, R12, R13, R14, R15.

V. Experiments

A. Corpus

The methodology of this research imposes certain requirements on a corpus used in it. Firstly, the corpus should contain a small amount of errors in its markup, because they make the error analysis of the method extremely complicated. This requirement makes unsuitable most of corpora with automatic markup or crowdsource-based markup. Secondly, every record in a corpus must contain only one sentence.

We examined the opportunity to use one of the available open Russian-language corpora. There are dozen open corpora available [17] but none of these meets all the aforementioned requirements. That is why we created a new one.

The corpus we created contains 1204 sentences extracted from hotel reviews from https://www.trivago.ru. The sentences were preprocessed:

- Smilies and emoticons were removed.
- Grammatical, spelling, and punctuation mistakes were fixed. It is required because mistakes can lead to incorrect PoS-tagging or syntax parsing.
- The *ё* letters, which are often substituted with *е* in Russian, were placed into their places. It is required to avoid PoS-tagging and parsing errors due to impossibility to distinguish some words without context after such a substitution, e.g., '*всё*' (*everything*) and '*все*' (*everybody*).

Two experts in linguistics marked each sentence as having positive, negative, or neutral sentiment. Sentences on which there was no agreement were excluded from the corpus.

It should be mentioned that the corpus does not contain sentences with a vague sentiment, e.g., *Кроме этого момента* *всё было хорошо* (*Aside from that, everything was all right*). This sentence contains two statements:

- The author was not satisfied with one certain aspect of his staying at the hotel.
- The author was satisfied with all the other aspects of his staying at the hotel.

The first statement expresses a positive emotion, and the second expresses a negative one. This means that the whole sentence sentiment is mixed and cannot be determined even by a human annotator, therefore it is unreasonable to apply the automatic approach on such sentences.

B. Sentiment dictionary

For the experiment we tried to use two sentiment dictionaries.

The first dictionary is RuSentiLex-2017 [18]. It consists of more than 12 thousand words and phrases, including slang words and swearings collected from Twitter.

Unfortunately, unlike the SentiWordNet sentiment dictionary, which was used in [9], RuSentilLex contains no sentiment scores. Instead it assigns to each presented word its type of sentiment (*positive*, *negative*, *neutral*, or *positive/negative*) and its source (*opinion*, *feeling*, or *fact*). To use this dictionary we developed a conversion algorithm that maps a sentiment and its source to the sentiment scores.

Firstly, we determine the neutral score. If word is not positive, negative, or positive/negative, i.e., it has no sentiment, we consider it totally neutral and assign 1.0 to its neutral score. Otherwise, the neutral score is assigned based on the source. The sentiment expressed by a feeling is less objective than the sentiment expressed by an opinion and the the sentiment expressed by an opinion is less objective than the opinion expressed by a fact:

$$obj = \begin{cases} 1.00, & \text{if sentiment is neutral,} \\ 0.50, & \text{if sentiment is not neutral and source is fact,} \\ 0.25, & \text{if sentiment is not neutral and source is opinion,} \\ 0.00, & \text{if sentiment is not neutral and source is feeling.} \end{cases}$$

When the neutral score is determined, we assign the difference between one and the neutral score to the positive score if the word sentiment is positive and to the negativity score if the word sentiment is negative. If the word sentiment is positive/negative, we divide that difference equally between the positivity and negativity scores.

The second sentiment dictionary we used is KartaSlovSent [19]. It contains more than 46 thousands words. Each word has four scores — positive, negative, neutral, and dunno. The first three scores represents shares of the vote for assigning positive, negative, and neutral sentiment to the word, whereas the fourth one represents the share of voters who failed to determine the sentiment of the word.

To use this dictionary we utilize these scores transformed in accordance to the recommendations given in [19]:

$$pos = \frac{positive\ share}{1 - dunno\ share},$$

TABLE I. SENTIMENT CLASSIFICATION PERFORMANCE ON A QUARTER OF THE CORPUS WITH RUSENTILLEX-2017 DICTIONARY

Class	Precision	Recall	F-score	No. of sentences
Positive	0.78	0.74	0.76	139
Neutral	0.32	0.82	0.46	57
Negative	0.86	0.18	0.30	105
Average	0.65	0.58	0.51	301
Weighted average	0.72	0.56	0.54	301

Accuracy = 0.56

TABLE II. SENTIMENT CLASSIFICATION PERFORMANCE ON A QUARTER OF THE CORPUS WITH KARTASLOVSENT DICTIONARY

Class	Precision	Recall	F-score	No. of sentences
Positive	0.65	0.87	0.74	139
Neutral	0.25	0.35	0.29	57
Negative	0.94	0.30	0.46	105
Average	0.61	0.51	0.50	301
Weighted average	0.68	0.57	0.56	301

Accuracy = 0.57

$$neg = \frac{negative\ share}{1 - dunno\ share},$$

$$obj = \frac{neutral\ share}{1 - dunno\ share}.$$

The words with the dunno share greater than 0.2 or having disagreement between the positive and negative scores greater than 0.05 are excluded.

C. Experiment results

In our experiments we initially used only on a quarter of the available corpus, then performed the error analysis and refined the approach according to its results. Finally, we measured the approach performance on the full corpus.

As the performance metrics we used both the simple average and the weighted average of precision, recall, and F-score. The simple average (also called arithmetic mean) is just the sum of the performance metrics for all the classes divided by the number of classes. The weighted average is the sum of the performance metrics for classes multiplied by the number of sentences in every class and divided by total number of sentences in the corpus. The reason of using the weighted average is the imbalance of classes in the corpus that may lead to incorrect assessment of the approach performance.

The results on a quarter of the corpus are shown in Table I and II.

From the experiment results it is clear that the classification performance is lower than the performance of the original approach [9]. Reduction is approximately 20% when compared to the movie reviews dataset results. This compasion is more correct than compasion to the Twitter results due to similarity of hotels and movies reviews.

It is also easy to notice that classification errors are distributed nonuniformly between the classes. The approach distinguish positive sentences from other sentiment classes rather well, but separation of negative and neutral classes is currently a drawback. The confusion matrices (Table III

TABLE III. SENTIMENT CLASSIFICATION CONFUSION MATRIX ON A QUARTER OF THE CORPUS WITH RUSENTILEX-2017 DICTIONARY

Actual \ Predicted	Positive	Neutral	Negative	Total
Positive	103	34	2	139
Neutral	9	47	1	57
Negative	20	66	19	105

TABLE IV. SENTIMENT CLASSIFICATION CONFUSION MATRIX ON A QUARTER OF THE CORPUS WITH KARTASLOVSENT DICTIONARY

Actual \ Predicted	Positive	Neutral	Negative	Total
Positive	121	17	1	139
Neutral	36	20	1	57
Negative	29	44	32	105

and IV) shows that the approach detects negative polarity very poor, classifying most of the negative sentences as neutral and almost quarter of the negative sentences as positive. The matrix also shows that the neutral sentences are incorrectly classified as positive rather than as negative.

It is clear from the confusion matrices that the sentiment dictionary has significant influence on the results. With RuSentiLex-2017 most of the errors are due to the poor negative and neutral classes distinction and low recall for the negative class. With KartaSlovSent the errors are connected with both poor distinction of negative and neutral, as well as positive and neutral classes. Generally, the results when using RuSentiLex-2017 are slightly better.

The classification metrics for the positive sentences are as good as in the original work. That gave us a hope that high performance of the sentiment analysis can be obtained after error analysis and refinement of the approach.

D. Error analysis

For error analysis and further approach refinement we chose to use the RuSentiLex dictionary only. The first advantage of that dictionary is the fact that it contains not only single words but also phrases that can be useful for handling idioms. The second advantage is that when using this dictionary, unlike KartaSlovSent, the only one main problem to be solved — negative and neutral sentences distinction.

To investigate the reasons of the performance decreasion, we collected information on the classification errors (Table V), i.e., we looked through all of the sentences our approach classified incorrectly and pointed out the errors.

We subdivided the errors into four groups:

- incorrect syntax tree parsing,
- incorrect sentiment bearers searching,
- imperfection of the rules,
- sentiment is beared by high-level sentence structure.

Incorrect syntax tree parsing, which is the most infrequent group of errors, includes the errors with part-of-speech tagging, dependencies parsing, or some other errors done by a syntactic parser. For example, in the phrase *внутренняя творческая сила человека* (*a human's inner creative force*) the parser incorrectly determined the word '*внутренняя*' as a dependant of '*творческая*', whereas in fact '*внутренняя*' is a dependant of '*сила*' (see Fig. 3). As a result, the phrase '*внутренняя сила*' is not determined as having the positive sentiment.

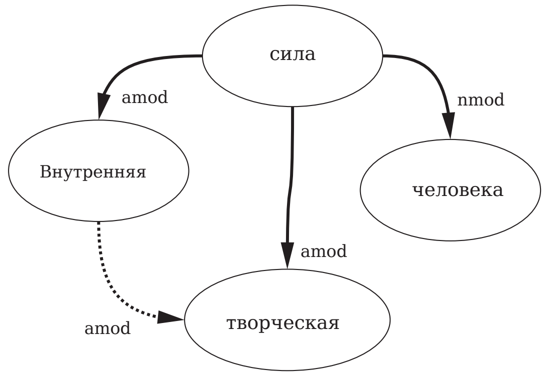

Fig. 3. Example of correct (solid lines) and incorrect (dotted lines) syntax tree parsing

The group of errors with the incorrect sentiment bearing words searching are divided into three subgroups:

- A word in its common meaning has a sentiment, but it is not present in the RuSentiLex sentiment dictionary. For example, in the sentence *Спасали от жары кондиционеры, но в большей части гостиницы была духота* (*The air conditioners helped a lot, but it was very stuffy in the most parts of the hotel*) the word '*духота*' has a negative sentiment, but there is no entry for it in RuSentiLex, therefore, it is assumed to be neutral.
- A word is incorrectly determined as positive or negative because its sentiment is context-dependent. For example, consider the sentence *Реальность соответствует фотографиям* (*The reality corresponds to the photos*). There are two records for '*реальность*' in RuSentiLex. One of them is considered as positive in the sense of feasibility and the other is neutral in the sense of objective reality. Due to insufficient context handling, the approach was unable to choose the correct record, which led to false assuming the sentiment of '*реальность*' as positive.
- A word has a sentiment only in the specific hotel reviews domain, hence, it is not present in the RuSentiLex dictionary. For example, the word '*слышимость*' in the sentence *В номерах фантастическая слышимость* (*Fantastic audibility in the rooms*) is a typical example of this kind of errors: '*слышимость*' is generally a neutral word but in the hotel room description its sentiment is negative.

Errors with the sentiment bearing words searching is the group of errors that occurs the most frequently, in more than 80% of the sentences. In some cases there are more than one errors

TABLE V. ERROR GROUPS ON A QUARTER OF THE CORPUS

Error	% of sentences
Incorrect syntax tree parsing	1.53
Incorrect sentiment bearers searching	80.15
Imperfection of the rules	15.27
Sentiment is beared by a high-level sentence structure	3.05
Total	100.00

with the sentiment bearing words searching in a one sentence.

The ‘imperfection of the rules’ group includes errors when application of one of the rules leads to incorrect sentiment identification. For example, in the sentence *Тренажёрка собственная маленькая, но хорошая сауна при ней* (*The own gym is small, but it has a good sauna*) there are two sentiment bearers — ‘*маленькая*’ and ‘*хорошая*’. Both of these words are important to determine the sentiment of the sentence, but application of R11 makes ‘*маленькая*’ to be assumed as having no sentiment.

The last group of errors is related to the situations when the sentiment of a sentence cannot be be determined using semantic rules because of the sentiment bearing by a high-level sentence structure and not particular sentiment bearer phrases. For example, the sentence *Брали номер с завтраком — при заселении сразу предупредили, что завтраков не будет* (*We booked a room with breakfast — at check-in we have been noticed that there would be no breakfasts*) is obviously negative, but there are no words with the negative sentiment in the sentence; the negative polarity is formed by a contradiction between *booked with breakfasts* and *there would be no breakfasts*.

We fixed the errors of syntax tree parsing by replacing the Taiga parser with SynTagRus. The improvement of sentiment bearers searching looked very important because of the vast majority of the errors occurred; it is described in the next section. The errors caused by the rules imperfection were not very frequent and we refused fixing them for now. Finally, we admitted that the proposed approach is principally unable to handle high-level sentence structure and the sentiments born by this structure and, hence, it is impossible to fix the errors caused by high-level sentence structure being inside its borders.

E. Improvements of sentiment bearers searching

The first improvement of sentiment bearers searching was made by using sentiment bearing pairs and triples of words from the sentiment dictionary. There are no word pairs and triples in SentiWordNet, that is why the original approach did not imply their using. We implemented searching sentiment bearing pairs and triples and achieved the increase of the average classification metrics by 3–4% (see Table VI). Such an increase was due to the progress in negative sentences detection. The performance scores of positive and neutral sentences distinction were slightly decreased probably because of false detection of previously unfound positive sentiment bearers in the neutral sentences.

TABLE VI. SENTIMENT CLASSIFICATION PERFORMANCE ON A QUARTER OF THE CORPUS AFTER IMPLEMENTING SENTIMENT BEARING PAIRS AND TRIPLES SEARCHING

Class	Precision	Recall	F-score	No. of sentences
Positive	0.75	0.77	0.76	139
Neutral	0.31	0.67	0.42	57
Negative	0.88	0.29	0.43	105
Average	0.56	0.57	0.54	301
Weighted average	0.71	0.58	0.58	301

Accuracy = 0.58

TABLE VII. SENTIMENT CLASSIFICATION PERFORMANCE ON A QUARTER OF THE CORPUS AFTER THE SENTIMENT DICTIONARY UPDATING

Class	Precision	Recall	F-score	No. of sentences
Positive	0.83	0.85	0.84	139
Neutral	0.40	0.72	0.51	57
Negative	0.93	0.49	0.64	105
Average	0.72	0.68	0.66	301
Weighted average	0.78	0.70	0.71	301

Accuracy = 0.70

The second improvement of sentiment bearers searching was related to sentiment dictionary fine tuning. From Subsection V-D it is clear that the RuSentiLex-2017 sentiment dictionary used in the experiments was imperfect that made a strong negative impact on the approach performance. This impact is not due to a drawback of the approach and does not allow to objectively assess its quality.

The most common issue of the dictionary is that many frequently used Russian adverbs having a strong sentiment like *хорошо* (*good*) and *плохо* (*bad*) are not present. Meanwhile, the adjectives from the same word families, such as *хороший* and *плохой* are present. Another issue is that many neutral words are considered as sentiment bearers, for example, *гроза* (*thunderstorm*) is considered as negative according to RuSentiLex.

That is why we examined all the sentiment bearing words in the corpus that were not present in the dictionary and added those of them that were domain-independent (i.e., suitable for a general purpose Russian sentiment dictionary) and having an undoubtful sentiment. We also excluded all the words that had doubtful or incorrect sentiment. All the performed modifications are given in Appendix.

Table VII shows that after the dictionary fine-tuning F_1-score increased by 12–13% and no performance metrics decreased. The most valuable impact was on the quality of negative and neutral sentences distinction.

F. Comparison with the state-of-the-art

To compare the proposed rule-based approach with the state-of-the-art, we firstly evaluated the rule-based approach and the BERT neural network on the full corpus. In the process of BERT training, 4-fold cross-validation was utilized. The results are presented in Tables VIII–XI. The rule-based approach on the full corpus performed slightly better than

TABLE VIII. SENTIMENT CLASSIFICATION PERFORMANCE ON THE FULL CORPUS

Class	Precision	Recall	F-score	No. of sentences
Positive	0.88	0.88	0.88	639
Neutral	0.48	0.75	0.58	232
Negative	0.94	0.57	0.71	333
Average	0.77	0.73	0.73	1204
Weighted average	0.82	0.77	0.78	1204

Accuracy = 0.77

TABLE IX. SENTIMENT CLASSIFICATION CONFUSION MATRIX ON THE FULL CORPUS

Actual \ Predicted	Positive	Neutral	Negative	Total
Positive	546	72	3	639
Neutral	50	173	9	232
Negative	24	118	191	333

on the quarter of the corpus due to the better distinction of negative and neutral sentences.

On averages, BERT shows slightly better (by 5%) results than the rule-based approach, particularly the rule-based approach falsely determines positive and negative sentences as neutral more often than BERT does. However, BERT falsely considers as negative more sentences that the rule-based approach. According to the other scores, the rule-based approach and BERT perform sentiment analysis almost equally well.

VI. CONCLUSION

We adapted the rule-based approach, which was proposed for the English language by Xie et al. and improved by Appel et al., for the Russian language. In order to take into account the highly flexible structure of a Russian sentence, we recreated the rules representation as algorithms over the syntax tree of a sentence. To analyze strengths and drawbacks of the adapted approach we evaluated it in on a quarter of a hotel reviews corpus; the F_1-measure of 0.51 was achieved.

As a result of the error analysis, we discovered that the vast majority of errors was caused by incorrect sentiment bearing words searching. We also discovered that the RuSentiLex-2017 sentiment dictionary we used for the experiment has very poor quality and does not contain many frequently used sentiment words, especially, adverbs. To measure the approach performance without the negative impact of the RuSentiLex quality we fine-tuned the sentiment dictionary by adding the missing sentiment bearing words. After the approach refinement the F_1-measure of 0.73 was achieved.

Comparison with BERT on the full hotel reviews corpus showed that the performance of the proposed approach is close to the state-of-the-art but only when a high-quality sentiment words dictionary is used. The rule-based approach distincts negative and neutral sentences slightly worse than BERT does. This gap can be probably filled by error analysis and further approach refinement, which are the main advantages of the proposed approach. Another important advantages include absence of the need of a large corpus to be trained on (unlike machine learning methods) and independence of the rules from any particular domain.

We must also notice that performance close to the state-of-the-art can already be achieved despite the fact that the currently used approach processes only a part of the clauses existing in Russian. There is a hope that introduction of new rules based on more detailed language study would improve the results.

The experiments showed that the most critical obstacle for application of rule-based approaches in Russian is the low quality of existing sentiment dictionaries. This pitfall should be fixed for further development of these approaches. As for rules extension, it looks perspective to improve word groups processing based on the clausal analysis of a sentence. The idea of propagating sentiment from dependants to their parent also looks promising. Another future research direction would include experiments on sentences from the other domains having other speech styles and more complex sentences.

TABLE X. BERT SENTIMENT CLASSIFICATION PERFORMANCE

Class	Precision	Recall	F-score	No. of sentences
Positive	0.91	0.92	0.91	639
Neutral	0.63	0.63	0.63	232
Negative	0.82	0.80	0.81	333
Average	0.78	0.78	0.78	1204
Weighted average	0.83	0.83	0.83	1204

Accuracy = 0.83

TABLE XI. `BERT SENTIMENT CLASSIFICATION CONFUSION MATRIX

Actual \ Predicted	Positive	Neutral	Negative	Total
Positive	588	36	15	639
Neutral	43	146	43	232
Negative	17	51	265	333

ACKNOWLEDGMENT

This work was supported by P.G. Demidov Yaroslavl State University Project No. VIP-016.

REFERENCES

[1] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 2005, pp. 347–354.

[2] L. K.-W. Tan, J.-C. Na, Y.-L. Theng, and K. Chang, "Sentence-level sentiment polarity classification using a linguistic approach," in *International Conference on Asian Digital Libraries*, 2011, pp. 77–87.

[3] M. S. Başarslan and F. Kayaalp, "Sentiment analysis on social media reviews datasets with deep learning approach," *Sakarya University Journal of Computer and Information Sciences*, vol. 4, no. 1, pp. 35–49, 2021.

[4] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-based systems*, vol. 89, pp. 14–46, 2015.

[5] S. Smetanin and M. Komarov, "Deep transfer learning baselines for sentiment analysis in Russian," *Information Processing & Management*, vol. 58, no. 3, pp. 102–121, 2021.

[6] O. Onyimadu, K. Nakata, T. Wilson, D. Macken, and K. Liu, "Towards sentiment analysis on parliamentary debates in Hansard," in *Joint international semantic technology conference*, 2013, pp. 48–50.

[7] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 45–63, 2014.
[8] Y. Xie, Z. Chen, K. Zhang, Y. Cheng, D. K. Honbo, A. Agrawal, and A. N. Choudhary, "MuSES: multilingual sentiment elicitation system for social media data," *IEEE Intelligent Systems*, vol. 29, no. 4, pp. 34–42, 2013.
[9] O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to the sentiment analysis problem at the sentence level," *Knowledge-Based Systems*, vol. 108, pp. 110–124, 2016.
[10] C. Gómez-Rodríguez, I. Alonso-Alonso, and D. Vilares, "How important is syntactic parsing accuracy? An empirical evaluation on rule-based sentiment analysis," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 2081–2097, 2019.
[11] M. A. M. Shaikh, H. Prendinger, and M. Ishizuka, "Sentiment assessment of text by analyzing linguistic features and contextual valence assignment," *Applied Artificial Intelligence*, vol. 22, no. 6, pp. 558–601, 2008.
[12] P. Panicheva, "ATEX: a rule-based sentiment analysis system processing texts in various topics," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialouge»*, vol. 2, 2013, pp. 101–112.
[13] X. Wu, S. Lv, L. Zang, J. Han, and S. Hu, "Conditional BERT contextual augmentation," in *International Conference on Computational Science*, 2019, pp. 84–95.
[14] L. Pivovarova, M. Du, and R. Yangarber, "Adapting the PULS event extraction framework to analyze Russian text," in *Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing*, 2013, pp. 100–109.
[15] J. Nivre, M.-C. de Marneffe, F. Ginter, Y. Goldberg, J. Hajič, C. D. Manning, R. McDonald, S. Petrov, S. Pyysalo, N. Silveira, R. Tsarfaty, and D. Zeman, "Universal Dependencies v1: A multilingual treebank collection," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, May 2016, pp. 1659–1666.
[16] A. Gvozdev, *The modern standard Russian language. Part II, Syntax*. Moscow: Prosveschevie, 1973, in Russian.
[17] S. Smetanin, "The applications of sentiment analysis for Russian language texts: Current challenges and future perspectives," *IEEE Access*, vol. 8, pp. 110 693–110 719, 2020.
[18] N. V. Loukachevitch and A. V. Levchick, "Creating a general Russian sentiment lexicon," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, May 2016, pp. 1171–1176.
[19] D. Kulagin, "Publicly available sentiment dictionary for the Russian language KartaSlovSent," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"*, vol. 16, 2021, pp. 19–23.

APPENDIX

RUSENTILEX-2017 CHANGES

20 words and word pairs that are considered to be positive: безопасно (opinion); идеально (opinion); любезно (opinion); слушаться (fact); уметь (fact); хочется вернуться (feeling); новый (fact); опрятно (opinion); отдохнуть (fact); отдых (fact); отлично (opinion); разнообразие (fact); разнообразный (fact); рекомендовать (opinion); спасибо (opinion); старинный (fact); супер (opinion); хорошо (opinion); шикарно (opinion); уникальный (opinion).

13 words and word pairs that were excluded from the dictionary: горячий; гроза; единственный минус; единственный недостаток; единственный плюс; критичный; подходить; продуктивность; реальность; светлый; сладкий; старый; стесняться.

64 words and word pairs that are considered to be negative: антисанитария (fact); бубнить (fact); врущий (fact); галдеть (fact); греметь (fact); громыхать (fact); далековато (opinion); драный (fact); ждать долго (opinion); забросить (fact); запах канализации (fact); застудить (fact); лишать смысла (opinion); ломать (fact); мешать (fact); не дождаться (fact); не дожидаться (fact); не открываться (fact); не первой свежести (opinion); невнимательно (fact); некомфортно (feeling); неочевидно (opinion); неприятно (feeling); никакой (adjective, opinion); обескуражить (feeling); обшарпать (fact); оставляет желать лучшего (opinion); отрицательное впечатление (feeling); отшибить (fact); очень тесно (opinion); плохо (opinion); поломан (fact); поломанный (fact); потерять (fact); пришлось подождать (fact); пришлось просить (fact); проблематично (fact); прокурить (fact); проходной двор (fact); развод (fact); разводить руками (feeling); разговаривать грубо (opinion); раздолбанный (fact); раздолбать (fact); разочароваться (feeling); разъезжаться (fact); сифонить (fact); сломан (fact); сломанный (fact); снизить оценку (fact); старенький (fact); странный (fact); сырость (fact); троечка (fact); угар (fact); упирать (fact); упираться (fact); фу (feeling); чёрствый (fact); чпокать (fact); ужасно (opinion); шепелявый (fact); шуметь (fact); шумно (fact).