Automatic Detection of Sentiment Towards Explicit Aspect in Russian Publicism Sentences Using Syntactic Structure

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

***Abstract*—The paper describes experiments on using the syntactic structure to improve the performance of sentiment detection towards an explicit aspect in Russian sentences. Three methods of syntactic structure usage were evaluated. The first is augmenting sentences from training set by changing their syntactic structure. The second is passing local context subtrees to Interactive Attention Network (IAN). The third is the novel semantic rule-based sentiment detection algorithm based on the previous authors' work. Sentences augmentation and passing local context did not exceeded the baseline, their macro F-scores were 0.68 and 0.69 respectively. The proposed rule-based algorithm performed almost as well as the baseline method, its macro F-score is 0.70. The ensemble of BERT-SPC and the rule-based algorithm performed with F-score 0.81, which exceeds the baseline by 10 %.**

I. Introduction

Automatic detection of sentiment towards an explicit aspect in a publicism sentence is a Natural Language Processing (NLP) task that can be formulated as follows: given a sentence and an aspect of the economic and social development, which is explictly mentioned in the sentence, it is required to determine the author's attitude to the aspect: whether it is positive, negative or neutral [1]–[4]. The aspect can be either an event, or a phenomenon, or a named entity — the only requirement is its explicit mention, i.e., it is called by one or more words in the sentence. This task is different from the sentence sentiment detection task (general sentiment detection task) because there could be more than one explicitly mentioned aspect in the sentence. For example, in the sentence *Fans of Manchester were going to make a mess but law enforcement retained order in the city* there is negative attitude towards the aspect *Fans of Manchester* and positive attitude towards the aspect *law enforcement*.

This paper is devoted to application of various approaches of using the syntactic structure to improve performance of detection of sentiment towards an explicit aspect in Russian publicism sentences. Because there is no strict word order in Russian language, we expect [5] that information about the syntactic structure of the sentence would improve detection quality for Russian publicism sentences, which is, in general, substantially lower than for English sentences at the moment [6], [7].

There are several approaches of using the syntactic structure for sentiment detection. It can be used directly, e.g., via a rule-based sentiment detection algorithm on a syntactic tree [5], [8]. The other approach is to utilize the knowledge on syntactic structure to improve the quality of an existing machine learning method of sentiment detection, e.g., using syntax knowledge to extract additional information to be passed to the method [9].

The rest of the paper is organized as follows. Section II describes the existing corpora in Russian and the methods previously used for detection of sentiment towards an explicit aspect. Experiments on improvement of existing neural network-based approaches with the use of syntactic information are described in Section III. Section IV describes the semantic rule-based sentiment detection algorithm, which uses the syntactic tree. The ensemble of this algorithm with the best of the considered neural network-based model as well as experiments with it are described in Section V. Conclusion summarizes the paper.

II. Related work

A. Corpora

Most of the recent works for Russian devoted to the detection of sentiment towards an explicit aspect in publicism sentences uses one of the two openly available corpora. The first one is CABSAR[1] created by Naumov et al. [3]. It consists of 6705 records, each containing a sentence, an aspect term and its position in the sentence (indices of start and end characters), and a mark of the sentiment towards the aspect: positive, negative, or neutral. Because there could be more than one aspect mentioned in the sentence, some records contains the same sentence but different aspect terms. The sentences were collected from different sources: LiveJournal (2105 records), Lenta.ru news agency (2603 records), and Twitter (1997 records). There are two types of aspects in CABSAR: person and organisation, i.e., all the aspects are named entities. The annotation was performed using Yandex.Toloka crowdsourcing platform. The agreement rate between annotators was 0.84 that indicates rather strong agreement. 34 % of the CABSAR records contains positive sentiment towards an aspect, 20 % contains negative sentiment, and 46 % contains no sentiment

[1] https://github.com/sag111/CABSAR

towards aspects. The corpus is split into train and test sets, containing 80 % and 20 % of all records respectively.

The second corpus openly available for Russian is RuSentNE-2023[2] created by Golubev et al. [10]. It consists of 9482 records of the same structure as in CABSAR. There are five types of aspects in the RuSentNE-2023: person, organization, profession, country, and nationality, i.e., all the aspects are named entities too. The sentences were collected from NEREL corpus that contains person-oriented news [11]. The annotation was performed by both students and experts in linguistics and later checked by a moderator. The agreement rate between annotators was 0.84 (the same as for CABSAR). 13 % of the RuSentNE-2023 records contains positive sentiment towards aspects, 15 % contains negative sentiment, and 72 % contains no sentiment towards aspects. It means that the class imbalance in RuSentNE-2023 is stronger than in CABSAR. RuSentNE-2023 is also split into train and test sets, containing 70 % and 30 % of records respectively. There is also set for final testing, but there are no sentiment labels published for it.

B. Methods

The only results on CABSAR were published by its creators [3]. The best F-score is 0.70 on the test set. It was achieved using IAN (Interactive Attention Network) and ELMo embedding vectors. IAN is the neural network architecture integrating two LSTM neural networks. One of LSTMs handles the aspect (its inputs are embedding vectors of the aspect term words) and the other handles the mention context (its inputs are embedding vectors of all the words of the sentence) [12]. IAN usually performs better on aspect-based sentiment analysis tasks than LSTM does because it can handle both the aspect, the mention context, and the relation between them [13], [14]. Naumov et al. also reported results of the sentiment lexicon-based algorithm evaluation on CABSAR. Its macro F-score is 0.31. This weak result can be explained by simplicity of the algorithm: it determines, which sentiment, positive or negative, is more common among the words of the sentence.

CABSAR was also used by Moloshnikov et al. [15], but only for training, and no performance scores were published, whereas these researchers achieved a rather high sentiment detection performance with IAN on RuSentNE-2023: macro F-score is 0.68 on the test set with RuBERT-large embeddings. In the same paper the authors also reported that the sentiment detection method based on text-to-text generative model ruT5-large outperforms IAN on the test set by 11 % but this approach a requires larger train set. The ruT5-large model in the paper was trained on union of RuSentNE-2023 train and test sets, and the CABSAR corpus. It makes the performance achieved with ruT5-large incompatible with the other results achieved on RuSentNE-2023.

The best reported results for RuSentNE-2023, with macro F-score between 0.75 and 0.78, were achieved by three researchers: Kabaev et al. [16], Glazkova [17], and Sanockin et al. [18]. All of them experimented with different embeddings and masking techniques for the basic BERT classification model called BERT-SPC (Sequence Pair Classification) [19].

It should be mentioned that at the moment there are no published results of more complex BERT-based models, such as Local Context Focus BERT (LCF-BERT) [20], evaluated on CABSAR or RuSentNE-2023.

Rusnachenko et al. [6] also published the results of different large language model-based methods evaluation on RuSentNE-2023. The best F-score of these methods in zero-shot mode is 0.64 on the test set achieved using GPT-4, which is 14 % lower than the best BERT-SPC result. The best F-score of fine-tuned models is 0.76 on the test set achieved with Flan-T5xl, which is 2 % lower than the best BERT-SPC result.

There are some rule-based methods of sentiment detection towards an explicit aspect for non-Russian languages. For example, Brunova et al. [4] proposed an algorithm for English based on the set of 15 sentiment detection rules. The rules handle different ways of sentiment expression: sentimental words, coordinating and adversative conjunctions, amplifying particles, and so on. The algorithm was evaluated on the set of 300 Reddit texts related to politics and reached accuracy of 0.92.

Bayraktar et al. [21] proposed a rule-based algorithm for Turkish. The algorithm is based on the sentiment dictionary SentiTurkNet and the set of semantic rules that handles 6 most popular Turkish conjunctions. The algorithm was evaluated on SemEval ABSA 2016 and reached accuracy 0.52. The error analysis performed by the authors showed that significant amount of errors is caused by poor sentiment dictionary quality and lack of semantic rules for specific ways of sentiment expression.

Unfortunately, there a is lack of rule-based methods of sentiment detection towards an explicit aspect for Russian publicism, there are rule-based methods of general sentiment detection for Russian publicism. In 2013 Panicheva proposed ATEX — rule-based sentiment analysis system for Russian texts [22]. The set of semantic rules used in ATEX handles different ways of sentiment expression in Russian publicism.

In our previous work [5] we proposed a novel rule-based general sentiment detection algorithm. The algorithm processes constituency tree of the sentence that represents its phrase structure: leafs represent words and internal nodes represent phrases. The algorithm detects sentiment of each phrase by sentiments of its constituent parts according to a set of semantic rules. The idea of the algorithm is quite similar to the idea of the algorithm of sentiment detection towards aspect described in [4].

III. Experiments with neural network-based models

A. Experiment design

The experiments in this research were conducted on records from the CABSAR corpus. This decision has three reasons. The first reason is that the annotation procedure of CABSAR

[2]https://github.com/dialogue-evaluation/RuSentNE-evaluation

is well-structured and more reliable than the annotation procedure of RuSentNE-2023. The second reason is that the class imbalance in CABSAR is weaker than in RuSentNE. The third reason is the presence of sentences from both blog records and news texts in CABSAR, which makes this corpus more relevant to this research.

The evaluation was done only on sentences extracted from LiveJournal and Lenta.ru, excluding the ones from Twitter, because tweets have their own specific speech style, which significantly differs from publicism [23]. Our experiments showed that performance on initial CABSAR and CABSAR without tweets (further referred just as CABSAR) differs no more than 1 %.

We chose BERT-SPC [19] and IAN [12] as baseline models because in previous research they showed themselves as the most efficient for the task under consideration [6], [10]. We also experimented with Bi-LSTM and LCF-BERT models [20].

As the goal of our research is using the syntactic structure to improve performance of sentiment detection towards an explicit aspect in Russian publicism sentences, we utilized three approaches: augmentation of sentences from the train set by changing their syntactic structure, passing local context subtrees of aspect terms to IAN, and semantic rule-based sentiment detection algorithm. Finally, we combined BERT-SPC and semantic rule-based in an ensemble.

During the experiments, we used Python programming language and PyTorch framework [24]. Constituency trees were constructed using the algorithm described in our previous work [25] and Stanza dependency parser [26].

All the performance metrics during evaluation were calculated on the test set of CABSAR (cross-validation was not used) to compare the achieved results with other published results on this dataset.

We used macro F-score because it is robust to the class imbalance present in the corpus, moreover it is used in the most of the published works.

B. Baseline

To determine the baseline for our own methods involving the syntactic structure, we evaluated four neural network-based methods of sentiment detection towards an explicit aspect: Bi-LSTM with ELMo embeddings, IAN with ELMo embeddings, BERT-SPC, and LCF-BERT. The best macro F-scores of the evaluated methods on the test set and hyperparameters, which was used to achieve best scores, are shown in Table I. All the models were trained until there was no macro F-score growth on the test set during 5 epochs.

We investigated the reason why the performance of the advanced LCF-BERT model was lower than the performance of the simpler BERT-SPC model. It showed that LCF-BERT suffers from overfitting: its performance on the train set grows too fast without simultaneous growth of the performance on the test set (see Fig. 1). It looks like the simpler BERT-SPC model works better when the the size of the train set is not very large and the classes are imbalanced: its performance increases

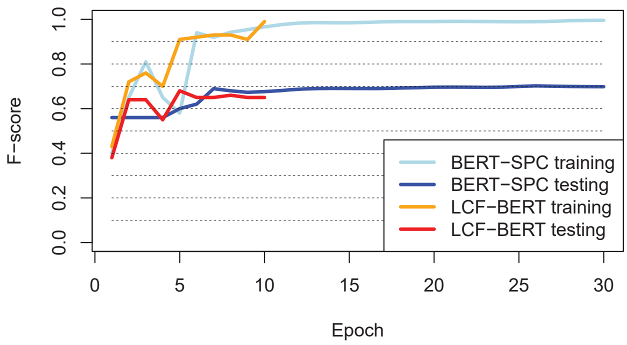

Fig. 1. BERT-SPC and LCF-BERT learning curves

slower than LCF-BERT performance does but it continues to increase when LCF-BERT performance stops to grow.

Performance metrics and the confusion matrix for BERT-SPC are shown in Tables II and III respectively. Performance for records with neutral sentiment is the highest, and performance for negative sentiment is the lowest. Positive and negative classes have similar precisions, but recall for the positive class is significantly higher than for the negative class.

C. Data augmentation

As neural neural network-based sentiment detection methods performs better when they are trained on larger datasets, we suppose that addition of more records with positive and negative sentiment into the train set would improve the sentiment detection performance. This can be achieved by the augmentation of the train set with additional sentences having modified syntactic structure.

The following augmentation techniques [27] were used:

- random syntactic groups order swap;
- random neutral syntactic groups and neutral words insertion;
- random replacement of sentimental words with synonyms with the same sentiment (according to RuWordNet [28] thesaurus and RuSentiLex [29] sentiment dictionary).

Using these techniques, the number of records with positive or negative sentiment in the extended corpus were doubled. However, the best macro F-score achieved by BERT-SPC trained on the extended corpus was 0.68, which is lower than the best BERT-SPC result for the initial train set. It was achieved with the same training parameters as for non-extended train set, but the training took 11 epochs, which is significantly lesser than the training on the non-extended set took. Performance metrics and the confusion matrix are shown in Tables IV and V respectively.

D. IAN with using context of aspect mention

The next experiment used the idea of extracting the words closely related to the aspect term in the sentence and passing

TABLE I. BASELINE METHODS EVALUATION RESULTS

Method	Macro F-score	Hyperparameters	No. of training epochs
Bi-LSTM	0.61	ELMo trained on Russian WMT News[3], hidden layers size 128, 3 recurrent layers, dropout rate 0.5, batch size 16, Adam optimizer, learning rate 0.0001, weight decay rate 0.0001	48
IAN	0.68	ELMo trained on Russian WMT News, hidden layers size 100, 1 recurrent layer in each LSTM, dropout rate 0.625, batch size 16, Adam optimizer, learning rate 0.00004, weight decay rate 0.0008	28
BERT-SPC	0.71	pretrained *rubert-base-cased-sentence*[4], hidden layer size 512, dropout rate 0.25, batch size 16, SGD optimizer, learning rate 0.005, weight decay rate 0.005	30
LCF-BERT	0.68	pretrained *rubert-base-cased-sentence*, semantic-relative distance 5, hidden layer size 512, dropout rate 0.625, batch size 16, SGD optimizer, learning rate 0.005, weight decay rate 0.005	11

them to IAN in addition to the aspect term and the full text of the sentence. We supposed that using a local context of the aspect term (namely words or phrases around the aspect term forming one subtree in the constituency tree) would improve the classification performance without causing overfitting. Further in the text we refer to this IAN modification as IAN-Context. To extract the phrase that forms such a context we created the following algorithm.

At the first stage the algorithm searches the constituency tree of the sentence for the node N that directly contains the aspect term. If the aspect term is a single word, N is the leaf representing the word itself. If the aspect term consists of two or more words, N is the root of the minimum subtree containing all the words of the aspect term. If there is more than one aspect term in the sentence, the algorithm uses the one specified in the record of the corpus.

At the second stage the algorithm searches for the d-th ancestor of N and extracts the subtree that has this ancestor as the root node. If there is no d-th ancestor (e.g., N itself is the root of the constituency tree), the maximum possible ancestor is used. The more is d, the broader context of the aspect term is extracted. We experimented with d equals 1, 2, and 3. All the words (leaf nodes) in the extracted subtree are combined into one phrase (a string) in the same order as in the sentence.

The best macro F-score of 0.69 was achieved for d equals 2 and the same training parameters as for base IAN. Performance metrics and confusion matrices for basic IAN and IAN-Context are shown in Tables VI and VII respectively.

There is a small performance increase for IAN-Context in comparison with IAN. This increase is due to the better negative sentiment detection: F-score for negative sentiment changed from 0.58 to 0.62 because IAN-Context detected more records as having negative sentiment than IAN did. But it also detected less records as having positive sentiment. The learning curves (see Fig. 2) show that IAN-Context overfits slower than IAN: its F-score on test set stops to grow slightly later than IAN. Therefore, the difference smoothes the effect of imbalanced classes, but IAN-Context still performs worse than BERT-based methods.

IV. SEMANTIC RULE-BASED ALGORITHM

A. Relation between sentiment towards an explicit aspect and general sentiment

Since there is a rather efficient rule-based algorithm for detection of the general sentence sentiment [5], it is natural to take it as a base for rule-based algorithm for sentiment detection towards an explicit aspect.

To estimate the relation between general sentiment and sentiment towards an explicit aspect, we performed the following experiment. The algorithm for detection of the general sentiment was executed on the CABSAR corpus and its results were treated as the sentiment towards an explicit aspect. Performance metrics and the confusion matrix are shown in Tables VIII and IX respectively.

The result of the experiment is very weak: macro F-score is 0.47, whereas the best macro F-score for the general sentiment detection algorithm in the article [5] is 0.76. It is not surprising and shows the essential differences between the two tasks.

TABLE II. BERT-SPC SENTIMENT DETECTION PERFORMANCE

Class	Precision	Recall	F-score	No. of sentences
Positive	0.68	0.73	0.70	252
Neutral	0.78	0.78	0.78	490
Negative	0.67	0.61	0.64	168
Macro average	0.71	0.70	0.71	910

Accuracy = 0.73

TABLE III. BERT-SPC SENTIMENT DETECTION CONFUSION MATRIX

Real \ Predicted	Positive	Neutral	Negative	Total
Positive	184	58	10	252
Neutral	69	380	41	490
Negative	19	47	102	168

However, this result is far better than random (according to [3] the result can be considered random if macro F-score is near 0.30). It shows the similarity of the tasks and corroborates the initial idea of using the developments of the general sentiment detection algorithm to propose an algorithm for detection of sentiment towards an explicit aspect.

B. Algorithm for detection of sentiment towards explicit aspect

The algorithm for detection of the general sentiment [5] assigns a sentiment to each node of the constituency tree during its execution. The algorithm for detection of sentiment towards explicit aspect, which we propose in this paper, uses this tree and the aspect term to detect the sentiment towards an explicit aspect. It is organized as follows.

At the first stage it searches the constituency tree for the node N that directly contains the aspect term. It is done the same way as in the first stage of the phrase extraction algorithm in Subsection III-D. If there is more than one aspect term in the sentence, the next stage is executed for all of them, then the results are aggregated according to the rule given below.

At the second stage the algorithm walks from N to the constituency tree root. For each node along this path it checks if any of the semantic rules described below is applicable. The rules are ordered by priority: if there is more than one applicable rule, the earliest in the list is used to detect the sentiment. If the root is reached and there are no applicable rules, the sentiment towards the aspect is considered neutral.

If there is more than one aspect term in the sentence, the previous stage results in multiple sentiment labels (one label for each aspect term). In this case the following sentiment aggregation rule is used. If all the labels are the same, their sentiment is the result of the algorithm. If at least one of the labels is positive and there is are no negative labels, the result is positive, and vice versa. If there are both positive and negative labels, the sentiment towards the aspect is considered neutral.

The algorithm uses the following 19 semantic rules representing the most common ways of sentiment expression towards explicit aspects in Russian elaborated with experts in Russian linguistics:

1) If the aspect term itself has a general sentiment (i.e., the sentiment of the corresponding syntactic group is different from neutral), then the result of the algorithm is this sentiment: if the word itself is sentimental, then by using it, the author expresses positive or negative attitude towards who or what is called by it.
2) If the aspect term or the word defined as related to the aspect is a subject, complement, or has a function of definition or determinative complement in a subject or complement group, and that group is neutral towards the aspect, the sentiment of the definition or determinative complement group is used to determine the sentiment towards the aspect: *Офицер подмосковной воинской части* ***Олег Леонтьев, обвиняемый в гибели солдата-срочника****, попросил в суде отправить его на службу в Сирию* (***Oleg Leontyev****, an officer of a military unit near Moscow,* ***accused in the death of an enlisted soldier****, asked in court to send him to serve in Syria*).
3) If the aspect term or the word defined as related to the aspect is a determinative complement and the complement group is neutral towards the aspect, the sentiment of the word being defined is used to determine the sentiment towards the aspect: *До* ***победы Цветкова*** *на пьедестал Кубка мира удавалось взобраться только Шипулину, который одержал победу в спринте на этапе Кубка мира в финском Контиолахти.* (*Before* ***Tsvetkov's victory****, only Shipulin had managed to climb the World Cup podium, winning the sprint at the World Cup stage in Kontiolahti, Finland*).
4) If the aspect term or the word defined as related to the aspect is a subject, or plays the role of a determiner or determinative complement in a subject group, and the

TABLE IV. SENTIMENT DETECTION PERFORMANCE OF BERT-SPC TRAINED ON THE EXTENDED CORPUS

Class	Precision	Recall	F-score	No. of sentences
Positive	0.67	0.70	0.68	252
Neutral	0.76	0.73	0.74	490
Negative	0.60	0.61	0.61	168
Macro average	0.67	0.68	0.68	910

Accuracy = 0.70

TABLE V. SENTIMENT DETECTION CONFUSION MATRIX OF BERT-SPC TRAINED ON THE EXTENDED CORPUS

Real \ Predicted	Positive	Neutral	Negative	Total
Positive	176	64	12	252
Neutral	75	358	57	490
Negative	13	52	103	168

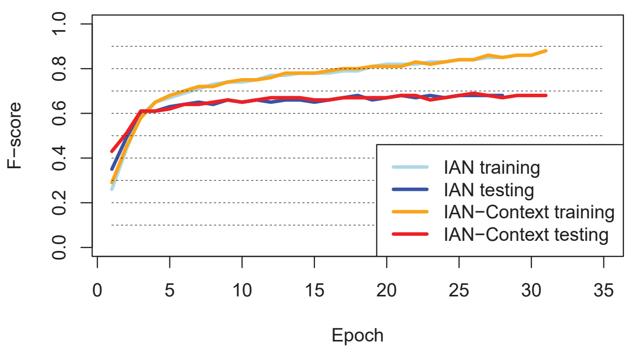

Fig. 2. IAN and IAN-Context learning curves

TABLE VI. IAN AND IAN-CONTEXT SENTIMENT DETECTION PERFORMANCE

Method	IAN			IAN-Context			No. of sentences
Class	Precision	Recall	F-score	Precision	Recall	F-score	
Positive	0.68	0.77	0.72	0.71	0.71	0.71	252
Neutral	0.75	0.70	0.73	0.75	0.75	0.75	490
Negative	0.58	0.58	0.58	0.62	0.61	0.62	168
Macro average	0.67	0.68	0.68	0.69	0.69	0.69	910

Accuracy of IAN = 0.70
Accuracy of IAN-Context = 0.71

TABLE VII. IAN AND IAN-CONTEXT SENTIMENT DETECTION CONFUSION MATRICES

Method	IAN			IAN-Context			Total
Actual \ Predicted	Positive	Neutral	Negative	Positive	Neutral	Negative	
Positive	193	54	5	178	65	9	252
Neutral	80	345	65	67	369	54	490
Negative	12	59	97	6	59	103	168

TABLE VIII. PERFORMANCE METRICS IF THE GENERAL SENTIMENT IS ASSUMED TO BE THE SENTIMENT TOWARDS AN EXPLICIT ASPECT

Class	Precision	Recall	F-score	No. of sentences
Positive	0.52	0.51	0.52	252
Neutral	0.59	0.31	0.41	490
Negative	0.34	0.80	0.47	168
Macro average	0.48	0.54	0.47	910

Accuracy = 0.46

TABLE IX. CONFUSION MATRIX IF THE GENERAL SENTIMENT IS ASSUMED TO BE THE SENTIMENT TOWARDS AN EXPLICIT ASPECT

Real \ Predicted	Positive	Neutral	Negative	Total
Positive	129	86	37	252
Neutral	107	154	229	490
Negative	10	23	135	168

subject group does not contain a positive or negative sentiment towards the aspect, then the predicate group is used to determine the sentiment towards the aspect: *Эта* ***книга Тиссо имела широчайший резонанс*** *в Европе и выдержала за полтора столетия 63 издания.* (*This* ***book by Tissot had the widest resonance*** *in Europe and went through 63 editions in a century and a half.*) If the author informs that an aspect or something closely related to an aspect plays a positive or negative role, the author is more likely to treat the aspect positively.

5) A rule similar to the previous one applies if the aspect term or the word defined as related to the aspect names the active person in a passive voice construction: *Духовная стрелка – это* ***просвещённая, чуткая, воспитанная Церковью совесть*** (*A spiritual arrow is an* ***enlightened, sensitive, conscience nurtured by the Church***).
6) If the aspect term or the word defined as related to the aspect acts as a direct complement in the syntactic group of the predicate with the semantics of desire, aspiration, preference, the sentiment towards the aspect is considered positive: *Удобность оформления кредита онлайн отметили 27 процентов опрошенных, такое же число респондентов отметили, что банк не готов выдать небольшую сумму на короткий срок,* ***поэтому предпочитают МФО*** (*Convenience of applying for a loan online was noted by 27 percent of respondents, the same number of respondents noted that the bank is not ready to give a small amount for a short period of time,* ***so they prefer MFOs***).
7) If the aspect term or the word defined as related to the aspect acts as an indirect complement in the syntactic group of the predicate with the semantics of expressing gratitude, the sentiment towards the aspect is considered positive: *Ковальчук* ***выразил благодарность руководству и болельщикам СКА*** *и заявил, что желает ближайшие 2–3 года "попытать свои силы в Америке"* (*Kovalchuk* ***expressed his gratitude to the SKA management and fans*** *and stated that he wants to "try his hand in America" in the next 2–3 years*).
8) If the aspect term or the word defined as related to the aspect is defined as having a quality in the highest degree, the sentiment towards the aspect is considered positive: ***Крупнейший в мире сайт о путешествиях TripAdvisor*** *признал «Аэрофлот» лучшей авиакомпанией Европы, а его бизнес-класс — лучшим в мире* (***The world's largest travel website TripAdvisor*** *recognized Aeroflot as the best airline in Europe and its business class as the best in the world*).
9) A rule similar to the previous one is applied if the aspect term or the word defined as related to the aspect

is defined as one superior to all others in some way: ***Ричард Прайор — один из самых знаменитых*** *американских комиков* (***Richard Pryor is one of the most famous*** *American comedians*).

10) If the aspect term or the word defined as related to the aspect is given by the author as an example of something positive or negative, the sentiment towards the aspect is considered positive or negative, respectively: ***На примере Филарета Гальчева*** *они могут увидеть, каких* ***успехов в бизнесе можно добиться*** (***On the example of Filaret Galchev*** *they can see* ***what success in business can be achieved***).
11) If the aspect term or the word defined as related to the aspect is defined as being protected from something negative, the sentiment towards the aspect is considered positive: *Иосиф в течение всех 15 лет старался* ***оградить*** *себя и* ***Икону от попыток использовать Её в своих корыстных целях*** *или манипулировать кого-то с Её помощью* (*Joseph tried for all 15 years to* ***shield*** *himself and* ***the Icon from attempts to use It for his own selfish purposes*** *or to manipulate someone with Its' help*).
12) If the aspect term or the word defined as related to the aspect acts as an indirect complement in a phrase with the semantics of expressing distrust or dissatisfaction, the sentiment towards the aspect is considered negative: *Как отмечается, духовенство во многом* ***недовольно*** *решениями* ***Дутерте****, особенно его кровавой борьбой с наркоторговцами* (*As noted, the clergy is largely* ***unhappy with Duterte's decisions****, especially his bloody fight against drug traffickers*).
13) If the aspect term or the word defined as related to the aspect acts as an indirect complement in a phrase with the semantics of getting something defined positively, the sentiment towards the aspect is considered positive: ***Приятные воспоминания*** *от этого события* ***остались у Кейт Уинслет*** *на всю жизнь* (***Pleasant memories*** *from this event* ***remained with Kate Winslet*** *for life*).
14) If the aspect term or the word defined as related to the aspect acts as an indirect complement in a phrase with the semantics of preference of something negative or neutral to something positive, the sentiment towards the aspect is considered negative: *Значит* ***для Алиева торговля нефтью важнее развития*** *малого и среднего бизнеса* (*So* ***for Aliyev oil trade is more important than development*** *of small and medium business*).
15) If the aspect term or the word defined as related to the aspect acts as a subject in a non-negative phrase with the semantics of doing something, the sentiment towards the aspect is considered positive: *Бывший кандидат в президенты России* ***Павел Грудинин*** *сбрил усы, тем самым* ***выполнив обязательство****, данное журналисту Юрию Дудю в рамках их спора* (*Former Russian presidential candidate* ***Pavel Grudinin*** *shaved off his mustache, thus* ***fulfilling a commitment*** *made to journalist Yury Dudy as part of their dispute*).
16) If the aspect term or the word defined as related to the aspect acts as an indirect complement in a phrase with the semantics of containing something positive or negative, the sentiment towards the aspect is considered positive or negative, respectively: ***В Российский клуб православных меценатов входят высокопрофессиональные юристы****, способные добиваться восстановления справедливости* (***In The Russian Club of Orthodox Patrons*** *we can find* ***highly professional lawyers*** *who are capable of seeking to restore justice*).
17) If the aspect term or the word defined as related to the aspect acts as a subject or indirect complement in a phrase with the semantics of lack of any positive or neutral qualities, the sentiment towards the aspect is considered negative: ***Дропбокс не имеет и половины сервисов*** *4shared* (***Dropbox does not have half*** *of 4shared's* ***services***).
18) If the aspect term or the word defined as related to the aspect acts as a subject in the main part of a compound sentence whose adjectival part reports a positively evaluated motive for action, the sentiment towards the aspect is considered positive: ***Массети*** *нарушил свою анонимность,* ***чтобы сделать крупное пожертвование на благотворительность*** (***Masseti*** *broke his anonymity* ***to make a large donation to charity***). If the author evaluates the motive of action positively, he thus expresses a positive attitude towards the subject of the action.
19) If the aspect term or the word defined as related to the aspect is used to compare it to something characterized positively, the sentiment towards the aspect is also considered positive: *Что это за любовь, если ты раскрываешь свою сокровенную мечту, весь такой* ***незащищенный, трогательный, как Бемби****, а в ответ: "Ты охренел?"* (*What kind of love is it if you reveal your innermost dream, and you're so* ***vulnerable and touching like Bambi*** *at this moment, and the response is, "Are you nuts?"*).

C. Algorithm evaluation

Performance metrics and the confusion matrix of the proposed algorithm of sentiment detection towards an explicit aspect are shown in Tables X and XI respectively. The algorithm detects positive and negative sentiment fairly well, but the quality of negative sentiment detection is significantly lower. The overall performance of the proposed algorithm is similar to the BERT-SPC performance: macro F-scores are 0.70 and 0.71 respectively. However, the rule-based algorithm confuses positive sentiment with negative less frequently than BERT-SPC does: the total numbers of such errors are 14 and 29 respectively.

V. Ensemble of BERT-SPC and rule-based algorithm

To improve the performance of sentiment detection towards an explicit aspect we combined two methods with the best performance, BERT-SPC and the rule-based algorithm from Section IV, in an ensemble. The ensemble uses outputs of the last BERT-SPC layer [30] and the logistic regression model described below to predict, if the result of BERT-SPC is correct or not. If the BERT-SPC result is correct, it is considered as the sentiment towards the explicit aspect; otherwise, the result of the recursive algorithm is used.

As the input data for the logistic regression the outputs of the last BERT-SPC layer were used. The training was done on the same train set as for the other models. The resultant regression equation is

$$p_{\text{true}} = \frac{e^{\alpha}}{1+e^{\alpha}},$$

$$\alpha = -8.064 + 2.265 \cdot x_{\max} + 0.651 \cdot x_{\text{pos}} + 1.097 \cdot x_{\text{neut}} + 1.074 \cdot x_{\text{neg}},$$

where

- p_{true} is the probability that the result of BERT-SPC is correct (the threshold value is 0.5);
- x_{pos}, x_{neut}, and x_{neg} are outputs of the last BERT-SPC layer for positive, neutral, and negative sentiment classes respectively;
- $x_{\max}$ is the maximum value of x_{pos}, x_{neut}, and x_{neg}.

F-score of the regression model is 0.70, therefore the model predicts BERT-SPC errors rather efficiently and can be used in the ensemble.

We also tried to use SVM and neural networks-based methods to predict errors of BERT-SPC, but did not succeeded due to imbalanced training data.

When using the ensemble, the sentiments of 64 % sentences were detected using BERT-SPC and for 36 % sentences the rule-based algorithm was used. Performance metrics and the confusion matrix are shown in Tables XII and XII respectively.

Macro F-score of the ensemble is 0.81, which is by 0.10 higher than F-scores of BERT-SPC and the rule-based algorithm. The ensemble detected neutral and positive sentiment fairly well, but the performance of negative sentiment detection is significantly lower. Recall of positive sentiment detection is 0.03 higher than precision, but for negative sentiment the balance is reverse: precision is 0.04 higher than the recall. 6 % of all negative sentences were detected as positive and only 2 % of all positive sentences were detected as negative. 9 % and 6 % neutral sentences were detected as positive and neutral sentiment respectively.

TABLE X. Performance of the proposed rule-based algorithm

Class	Precision	Recall	F-score	No. of sentences
Positive	0.69	0.73	0.71	252
Neutral	0.76	0.70	0.73	490
Negative	0.60	0.68	0.64	168
Macro average	0.69	0.71	0.70	910

Accuracy = 0.71

TABLE XI. Confusion matrix of the proposed rule-based algorithm

Real \ Predicted	Positive	Neutral	Negative	Total
Positive	185	62	5	252
Neutral	73	345	72	490
Negative	9	44	115	168

TABLE XII. Performance of the ensemble of BERT-SPC and the proposed rule-based algorithm

Class	Precision	Recall	F-score	No. of sentences
Positive	0.80	0.83	0.81	252
Neutral	0.85	0.85	0.85	490
Negative	0.78	0.74	0.76	168
Macro average	0.81	0.80	0.81	910

Accuracy = 0.82

TABLE XIII. Confusion matrix of the ensemble of BERT-SPC and the proposed rule-based algorithm

Real \ Predicted	Positive	Neutral	Negative	Total
Positive	208	39	5	252
Neutral	42	417	31	490
Negative	10	34	124	168

VI. Conclusion

This paper is devoted to the task of sentiment detection towards an explicit aspect. Its main goal is to improve the performance of sentiment detection by using the syntactic structure of a sentence. All the experiments were done on CABSAR sentences extracted from LiveJournal and Lenta.ru.

To estimate the baseline performance, we evaluated four neural network-based models: Bi-LSTM, IAN, BERT-SPC, and LCF-BERT. BERT-SPC showed the best performance, its macro F-score was 0.71.

In the work we utilized three approaches of using the syntactic structure to improve performance of sentiment detection towards an explicit aspect: augmentation of sentences from the train set by changing their syntactic structure, passing local context subtrees of aspect terms to IAN, and semantic rule-based sentiment detection algorithm.

Three augmentation methods (syntactic groups order swap, neutral words and syntactic groups insertion, and random replacement of sentimental words with synonyms) were used to extend the train set. However, the performance was not improved: the best macro F-score achieved by BERT-SPC trained on the extended set was 0.68.

In experiments with passing the local context of the aspect term to IAN the context was extracted by finding a subtree of the constituency tree that contains the aspect term. This allowed to improve the performance only by 0.01.

Then we proposed a rule-based algorithm of sentiment detection towards an explicit aspect. It uses the constituency tree, which nodes are assigned with the sentiment determined by the general sentiment detection algorithm, and 19 semantic rules representing the most common ways of sentiment expression towards an explicit aspect in Russian publicism. Macro F-score of the proposed algorithm was 0.70, which is close to F-score of BERT-SPC.

Finally, we combined BERT-SPC and semantic rule-based algorithm in an ensemble. It uses outputs of the last BERT-SPC layer and the logistic regression model. Macro F-score of the ensemble is 0.81, which is significantly higher than F-scores of BERT-SPC and rule-based algorithm used separately.

The achieved results can be interpreted as follows. When the training set is rather small, neural networks cannot learn many patterns existing in the language, but they learn patterns specific to the training set and its domain. On the other hand, rule-based algorithms rely on general sentiment expression patterns existing in the language. As a result, combining a neural network and a semantic rule-based algorithm in an ensemble can significantly improve the performance of sentiment detection.

ACKNOWLEDGMENT

The reported study was funded by the grant of Russian Science Foundation No. 23-21-00495 (https://rscf.ru/en/project/23-21-00495/).

REFERENCES

[1] U. Sirisha and S. C. Bolem, "Aspect based sentiment & emotion analysis with ROBERTa, LSTM," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 11, pp. 766–774, 2022.

[2] O. Alqaryouti *et al.*, "Aspect-based sentiment analysis using smart government review data," *Applied Computing and Informatics*, vol. 20, no. 1/2, pp. 142–161, 2024.

[3] A. Naumov *et al.*, "Neural-network method for determining text author's sentiment to an aspect specified by the named entity," in *CEUR Workshop Proceedings*, vol. 2648, 2020, pp. 134–143.

[4] E. Brunova, Y. Bidulya, and A. Gorbunov, "Aspect-based sentiment analysis of political discourse," *Tyumen State University Herald*, vol. 7, no. 3 (27), pp. 6–22, 2021, in Russian.

[5] A. Y. Poletaev, I. V. Paramonov, and E. I. Boychuk, "Semantic rule-based sentiment detection algorithm for Russian publicism sentences," *Modeling and Analysis of Information Systems*, vol. 30, no. 4, pp. 394–417, 2023, in Russian.

[6] N. Rusnachenko, A. Golubev, and N. Loukachevitch, "Large language models in targeted sentiment analysis," 2024. [Online]. Available: https://arxiv.org/abs/2404.12342

[7] S. Smetanin and M. Komarov, "Deep transfer learning baselines for sentiment analysis in Russian," *Information Processing & Management*, vol. 58, no. 3, p. 102484, 2021.

[8] S. Rani and P. Kumar, "Aspect-based sentiment analysis using dependency parsing," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 3, pp. 1–19, 2021.

[9] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020, pp. 3211–3220.

[10] A. Golubev, N. Rusnachenko, and N. Loukachevitch, "RuSentNE-2023: Evaluating entity-oriented sentiment analysis on Russian news texts," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual conference "Dialogue" (2023)*, vol. 22, 2023, pp. 130–141.

[11] N. Loukachevitch *et al.*, "NEREL: a Russian information extraction dataset with rich annotation for nested entities, relations, and wikidata entity links," *Language Resources and Evaluation*, vol. 58, pp. 547–583, 2024.

[12] D. Ma *et al.*, "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017, pp. 4068—4074.

[13] X. Liu *et al.*, "Multi-level attentional network for aspect-based sentiment analysis," in *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, 2020, pp. 17–22.

[14] A. Das and W. E. Zhang, "ABSA-Bench: Towards the unified evaluation of aspect-based sentiment analysis research," in *Proceedings of the 18th Annual Workshop of the Australasian Language Technology Association*, 2020, pp. 65–71.

[15] I. Moloshnikov *et al.*, "Named entity-oriented sentiment analysis with text2text generation approach," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual conference "Dialogue" (2023)*, vol. 22, 2023, pp. 361–370.

[16] A. Kabaev *et al.*, "HAlf-MAsked Model for named entity sentiment analysis," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual conference "Dialogue" (2023)*, vol. 22, 2023, pp. 433–441.

[17] A. Glazkova, "Fine-tuning text classification models for named entity oriented sentiment analysis of Russian texts," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual conference "Dialogue" (2023)*, vol. 22, 2023, pp. 104–116.

[18] L. Sanochkin *et al.*, "Simple yet effective named entity oriented sentiment analysis," in *Computational Linguistics and Intellectual Technologies: Papers from the Annual conference "Dialogue" (2023)*, vol. 22, 2023, pp. 459–468.

[19] Y. Song *et al.*, "Targeted sentiment classification with attentional encoder network," in *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series*. Springer International Publishing, 2019, p. 93–103.

[20] B. Zeng *et al.*, "LCF: A local context focus mechanism for aspect-based sentiment classification," *Applied Sciences*, vol. 9, no. 16, p. 3389, 2019.

[21] K. Bayraktar, U. Yavanoglu, and A. Ozbilen, "A rule-based holistic approach for Turkish aspect-based sentiment analysis," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 2154–2158.

[22] P. V. Panicheva, "ATEX: a rule-based sentiment analysis system processing texts in various topics," *Computational Linguistics and Intellectual Technologies*, vol. 2, no. 12, pp. 101–113, 2013, in Russian.

[23] A. Karysheva, "Lexical and grammatical features of Russian-language tweets in comparison with everyday spoken Russian," in *Proceedings of the 34th Conference of FRUCT Association*. IEEE, 2023, pp. 57–65.

[24] J. Ansel *et al.*, "PyTorch 2: Faster machine learning through dynamic Python bytecode transformation and graph compilation," in *ASPLOC'24: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. ACM, 2024, pp. 929–947.

[25] A. Y. Poletaev, I. V. Paramonov, and E. I. Boychuk, "Algorithm of constituency tree from dependency tree construction for a Russian-language sentence," *Informatics and Automation*, vol. 22, no. 6, pp. 1323–1353, 2023, in Russian.

[26] P. Qi *et al.*, "Stanza: A Python natural language processing toolkit for many human languages," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, 2020, pp. 101–108.

[27] M. Bayer, M.-A. Kaufhold, and C. Reuter, "A survey on data augmentation for text classification," *ACM Computing Surveys*, vol. 55, no. 7, pp. 1–39, 2022.

[28] N. Loukachevitch *et al.*, "Creating Russian wordnet by conversion," in *Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue" (2016)*, vol. 15, 2016, pp. 405–415.

[29] N. Loukachevitch and A. Levchik, "Creating a general Russian sentiment lexicon," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 1171–1176.

[30] J. Moon *et al.*, "Confidence-aware learning for deep neural networks," in *Proceedings of the 37th International Conference on Machine Learning*. JMLR.org, 2020, pp. 7034–7044.