arXiv:2111.10097v1 [cs.CL] 19 Nov 2021

Lexicon-based Methods vs. BERT for Text Sentiment Analysis

Anastasia Kotelnikova[1][0000−0001−9942−680X], Danil Paschenko[1][0000−0003−2671−8208], Klavdiya Bochenina[2][0000−0001−6025−0552], and Evgeny Kotelnikov[1,2][0000−0001−9745−1489]

[1] Vyatka State University, Kirov, Russia
kotelnikova.av@gmail.com
[2] ITMO University, Saint-Petersburg, Russia

Abstract. The performance of sentiment analysis methods has greatly increased in recent years. This is due to the use of various models based on the Transformer architecture, in particular BERT. However, deep neural network models are difficult to train and poorly interpretable. An alternative approach is rule-based methods using sentiment lexicons. They are fast, require no training, and are well interpreted. But recently, due to the widespread use of deep learning, lexicon-based methods have receded into the background. The purpose of the article is to study the performance of the SO-CAL and SentiStrength lexicon-based methods, adapted for the Russian language. We have tested these methods, as well as the RuBERT neural network model, on 16 text corpora and have analyzed their results. RuBERT outperforms both lexicon-based methods on average, but SO-CAL surpasses RuBERT for four corpora out of 16.

Keywords: Sentiment Analysis · Sentiment Lexicons · SO-CAL · SentiStrength · BERT

1 Introduction

The performance of sentiment analysis has improved dramatically[1] over the past few years, for example:

- the English-language corpus SST-5 (Stanford Sentiment Treebank – 5 classes), the accuracy increased from 45.70 (RNTN model [28]) to 59.10 (RoBERTa-large+Self-Explaining [29]);
- for the English-language corpus Yelp Reviews (5 classes), the error decreased from 37.95 (Char-level CNN model [36]) to 27.05 (XLNet model [35]);
- for the Russian-language news corpus ROMIP-2012 (3 classes) [6] F1-score increased from 62.10 (lexicon-based Polyarnik system) [17] to 72.69 [9].

[1] See for example: https://paperswithcode.com/task/sentiment-analysis.

This improvement in performance is mainly associated with the development of deep learning methods, especially various models based on the Transformer architecture [34], in particular, BERT [8].

However, deep neural network models are difficult to train: they need large amounts of data; training requires powerful expensive video cards with a large memory size; the learning process is time consuming and energy intensive [18]. Another issue is the complexity of interpreting the results of the models [1].

An alternative are rule-based (or lexicon-based) methods using sentiment lexicons [30]. They are fast, do not require training and are well interpreted [2]. But recently, due to the widespread use of deep learning, lexicon-based methods have receded into the background.

For the Russian language, there are several recent studies of deep learning models for sentiment analysis [9, 27]. However, there are currently no studies devoted to comparing lexicon-based methods and deep learning models.

We strive to close this gap and compare the fine-tuned deep neural network model RuBERT [16] with two lexicon-based methods adapted for the Russian language – SO-CAL [31] and SentiStrength [32]. For testing we have used 16 Russian-language text corpora, labelled by sentiment into 3 classes.

The contribution of this article is as follows:

- lexicon-based methods SO-CAL and SentiStrength have been adapted for the Russian language;
- performance evaluation of lexicon-based methods and RuBERT has been carried out for 16 Russian-language text corpora;
- the performance of SO-CAL and SentiStrength for 17 sentiment lexicons have been estimated: 9 publicly available Russian sentiment lexicons and a set of 8 combined lexicons;
- the classification results of lexicon-based methods and RuBERT have been analyzed.

2 Lexicon-based Methods and Tools

There are several tools for sentiment analysis on the base of sentiment lexicons:

- open source: SO-CAL [31], VADER [10], Pattern, TextBlob;
- proprietary: SentiStrength [32], SentText [25].

Taboada et al. developed SO-CAL[2] (Semantic Orientation CALculator) – a method and tool for determining the sentiment of texts in English and Spanish [31]. The sentiment is recognized on the basis of counting the weights of the sentiment words included in the text (only nouns, adjectives, verbs and adverbs are taken into account). A system of rules is also involved to account for the influence of lexical markers such as modifiers, negations, and irrealis markers. *Modifiers* are lexical markers that increase (e.g., *very*, *the most*) or decrease

[2] https://github.com/sfu-discourse-lab/SO-CAL.

(e.g., *slightly*, *somewhat*) the intensity of the next sentiment word. *Negations* (e.g., *not*, *nothing*) either invert the polarity of the next sentiment word, or shift its intensity towards the opposite polarity (for example, in SO-CAL, a shift is used, and in VADER – an inversion with a certain coefficient). *Irrealis markers* indicate that the sentiment score for a given sentence should not be taken into account. These markers are modal verbs (e.g., *could*, *should*), conditional words (e.g., *if*), some verbs (e.g., *expect*, *doubt*), a question mark, and quoted words.

Hutto and Gilbert proposed VADER[3] (Valence Aware Dictionary for sEntiment Reasoning) – a lexicon, method and tool for sentiment analysis of English texts [10]. The sentiment lexicon was built from the well-known dictionaries LIWC, ANEW, and General Inquirer and then crowdsourced in sentiment intensity. Also, emoticons, acronyms and slang were included into the lexicon. The VADER takes into account exclamation marks, capitalization, modifiers, negations and contrasts.

Pattern[4] is a web mining library that supports sentiment analysis in English and French [7]. For this, a lexicon of sentiment adjectives is used, often found in product reviews.

TextBlob[5] is a text processing library that includes two components for sentiment analysis – based on a naive Bayesian classifier and an implementation from the *Pattern* library.

Thelwall et al. developed SentiStrength[6], a tool for sentiment analysis of short social media texts based on the method bearing the same name [32]. The tool gives two scores for the input text: a negative score from –1 to –5 and a positive score from +1 to +5. The decision is based on a list of sentiment words with weights corresponding to the sentiment intensity. The method also takes into account modifiers, negations, question words, slang, idioms and emoticons.

Schmidt et al. developed SentText[7], a web-based sentiment analysis tool in the digital humanities [25]. The original version of SentText was developed for the German language using the SentiWS and BAWL-R dictionaries. This tool takes negations into account. SentText has the ability to visualize the results, including highlighting the sentiment words, information about the polarity of individual words and the text as a whole, as well as comparing texts by sentiment.

Of these tools, as far as we know, only SentiStrength has two adaptations for the Russian language[8], but we could not find a detailed description of their implementations.

In our work, we have adapted SO-CAL and SentiStrength for the Russian language. SO-CAL is the most advanced open source sentiment analysis tools. SentiStrength, despite being proprietary software, makes it easy to adapt to a new language. For this purpose, it is necessary to provide it with a sentiment

[3] https://github.com/cjhutto/vaderSentiment.
[4] https://github.com/clips/pattern.
[5] https://textblob.readthedocs.io.
[6] http://sentistrength.wlv.ac.uk.
[7] https://thomasschmidtur.pythonanywhere.com.
[8] Given on the website: http://sentistrength.wlv.ac.uk.

lexicon in the target language and other linguistic resources: lists of modifiers, negations, interrogative words, slang and idioms.

3 Materials and Methods

3.1 Lexicon-based Methods Adaptation

The adaptation of SO-CAL and SentiStrength to the Russian language includes the following steps:

- morphological analysis of input texts based on RNNMorph[9];
- preparation of a Russian-language sentiment lexicon. The existing lexicons were used to form it and are described in Subsection 3.2. A peculiarity of SO-CAL is to take into account only nouns, adjectives, verbs, adverbs;
- preparation of Russian-language lists of modifiers (e.g., *очень*, *едва*, *значительно*) and negations (e.g., *не*, *без*, *невозможно*). They were obtained by translating the corresponding SO-CAL and SentiStrength lists, as well as by adding Russian-language synonyms;
- preparation for SO-CAL of Russian-language lists of irrealis markers (e.g., *ожидать*, *можно*, *кто-нибудь*);
- modification of the SO-CAL source code for processing texts with the results of Russian morphological analysis;
- organization of programming interface with the desktop version of SentiStrength to submit input texts and process its results.

3.2 Sentiment Lexicons

The key resource for the considered sentiment analysis methods is the sentiment lexicon. The performance of sentiment analysis depends on the completeness and accuracy of such a lexicon. We have used 9 publicly available Russian sentiment lexicons (two of them – EmoLex and Chen-Skiena's – are Russian versions of multi-lingual lexicons) [13].

Each lexicon has been processed as follows:

- neutral words have been removed (if such were present in the lexicon);
- words that are both positive and negative in the lexicon have been removed (including the analysis of words with the spelling "е" and "ё");
- words containing Latin letters have been removed;
- all words have been converted to a lower case;
- words have been normalized using RNNMorph;
- only one occurrence of each element has been left (an element can be a separate word or phrase).

The characteristics of the lexicons are shown in Table 1.

[9] https://github.com/IlyaGusev/rnnmorph.

Table 1. The characteristics of sentiment lexicons.

Lexicon	Total	Positive elements		Negative elements	
		#	%	#	%
RuSentiLex [20]	12,560	3,258	25.9%	9,302	74.1%
Word Map [15]	11,237	4,491	40.0%	6,746	60.0%
SentiRusColl [14]	6,538	3.981	60.9%	2,557	39.1%
EmoLex [22]	4,600	1,982	43.1%	2,618	56.9%
LinisCrowd [11]	3,986	1,126	28.2%	2,860	71.8%
Blinov's lexicon [3]	3,524	1,611	45.7%	1,913	54.3%
Kotelnikov's lexicon [12]	3,206	1,028	32.1%	2,178	67.9%
Chen-Skiena's lexicon [4]	2,604	1,139	43.7%	1,465	56.3%
Tutubalina's lexicon [33]	2,442	1,032	42.3%	1,410	57.7%

We also used the "voting" procedure of these lexicons to build a set of 8 combined lexicons *Lex1..Lex8*: only those words that are included in at least N sentiment lexicons are included in the *LexN* lexicon. Thus, *Lex1* includes all sentiment words that occur in at least one lexicon. *Lex9* turned out to be empty – not a single item is included in all sentiment lexicons at the same time. The characteristics of the combined lexicons are shown in Table 2.

3.3 Text Corpora

For evaluation, we have used 16 public text corpora labelled by sentiment [26] (see Table 3), including:

- corpora of reviews about books, movies and cameras, as well as news articles from the ROMIP 2011 [5] and ROMIP 2012 [6] seminars;
- corpora of reviews about cars and restaurants, as well as tweets about banks and telecom companies of the SentiRuEval 2015 [21], SentiRuEval 2016 [19] and SemEval 2016 [23] seminars;
- the RuSentiment corpus containing posts on VKontakte [24];
- LinisCrowd corpus, including posts and comments from LiveJournal [11]. We have used texts labelled by one annotator as training data, and texts labelled by several annotators as test data.

4 Results

4.1 Experimental Setup

We have tested two lexicon-based methods of sentiment analysis adapted for the Russian language – SO-CAL and SentiStrength, as well as a deep neural network model RuBERT [16].

In general, lexicon-based methods can be used for sentiment analysis without training. However, since training corpora are available in our experiments, we have used them to tune the hyperparameters of the lexicon-based methods: we

Table 2. The characteristics of the combined sentiment lexicons.

Lexicon	Total	Positive elements		Negative elements	
		#	%	#	%
Lex1	33,080	13,443	40.6%	19,637	59.4%
Lex2	9,377	3,147	33.6%	6,230	66.4%
Lex3	4,325	1,521	35.2%	2,804	64.8%
Lex4	2,313	823	35.6%	1,490	64.4%
Lex5	1,266	475	37.5%	791	62.5%
Lex6	607	258	42.5%	349	57.5%
Lex7	240	114	47.5%	126	52.5%
Lex8	52	31	59.6%	21	40.4%

have chosen the optimal sentiment lexicons for both methods (out of 17 lexicons – see Subsection 3.2) and have determined the thresholds for positive and negative sentiment classes.

The thresholds are defined as follows. SO-CAL returns the single sentiment score s for the text to be converted to a class label (positive, negative, or neutral). We have fit two thresholds on the training data – t_{pos} and t_{neg}. The decision about the sentiment of the text c is made on the basis of the following expression:

$$c = \begin{cases} neutral,\ if\ s < t_{pos}\ and\ s > t_{neg}, \\ positive,\ if\ s \geq t_{pos}, \\ negative,\ if\ s \leq t_{neg}. \end{cases}$$

SentiStrength returns two sentiment scores for the text – positive s_{pos} and negative s_{neg}. We select two coefficients k_{neut} and k such that:

$$c = \begin{cases} neutral,\ if\ s_{pos} \leq k_{neut}\ and\ s_{neg} \leq k_{neut}, \\ positive,\ if\ s_{pos} > ks_{neg}, \\ negative,\ otherwise. \end{cases}$$

The results of selecting lexicons and threshold values are given in Subsection 4.2.

The pretrained RuBERT model was fine-tuned separately on each training corpus with the following hyperparameters: learning rate $2 \cdot 10^{-5}$, number of epochs 5, batch size 12. The results are given on average for five runs to reduce the influence of random weight initialization. The training has been carried out using the Google Colab Pro service on NVIDIA Tesla P100 and V100 video cards.

For all the corpora a three-class problem of sentiment analysis was solved – the classification of texts into positive, negative and neutral. We used the macro F1-score as the main performance metric.

4.2 Results of Experiments

We have run two series of experiments. In the first series, the training data were used to select the optimal hyperparameters for lexicon-based methods: lexicon

Table 3. Corpora characteristics.

Corpus	Type	Split	Total	Positive	Negative	Neutral
LinisCrowd	posts	train	28,853	7.7%	42.5%	49.8%
		test	14,260	9.5%	47.3%	43.2%
Romip 2011	book reviews	train	22,098	79.7%	9.3%	11.0%
		test	228	64.0%	6.2%	29.8%
	movie reviews	train	14,808	70.6%	12.7%	16.7%
		test	263	70.3%	10.7%	19.0%
	camera reviews	train	9,460	80.5%	10.6%	8.9%
		test	207	61.8%	17.9%	20.3%
Romip 2012	book reviews	test	129	77.5%	7.0%	15.5%
	movie reviews	test	408	65.2%	15.4%	19.4%
	camera reviews	test	411	85.4%	1.7%	12.9%
	news	train	4,260	26.2%	43.7%	30.1%
		test	4,573	31.7%	41.3%	27.0%
SentiRuEval 2015	car reviews	train	203	56.6%	14.8%	28.6%
		test	200	49.0%	13.0%	38.0%
	restaurant reviews	train	200	68.0%	14.0%	18.0%
		test	203	71.9%	12.8%	15.3%
	bank tweets	train	4,883	7.2%	21.7%	71.1%
		test	4,534	7.6%	14.4%	78.0%
	telecom tweets	train	4,839	18.8%	32.7%	48.5%
		test	3,774	9.1%	22.4%	68.5%
SentiRuEval 2016	bank tweets	test	3,302	9.1%	23.1%	67.8%
	telecom tweets	test	2,198	8.3%	45.8%	45.9%
SemEval 2016	restaurant reviews	test	103	67.0%	14.6%	18.4%
RuSentiment	posts	train	24,124	38.0%	15.2%	46.8%
		test	2,621	36.0%	9.8%	54.2%

and threshold values (see Subsection 4.1). In the second series the lexicon-based methods with selected hyperparameters were compared on test data with the fine-tuned RuBERT model.

The results of the first series of experiments on selecting the optimal sentiment lexicon are shown in Fig. 1. Kotelnikov's lexicon has turned out to be the best lexicon for SO-CAL, *Lex1* and *Lex2* – for SentiStrength (*Lex2* was used as the optimal lexicon, being the smaller one).

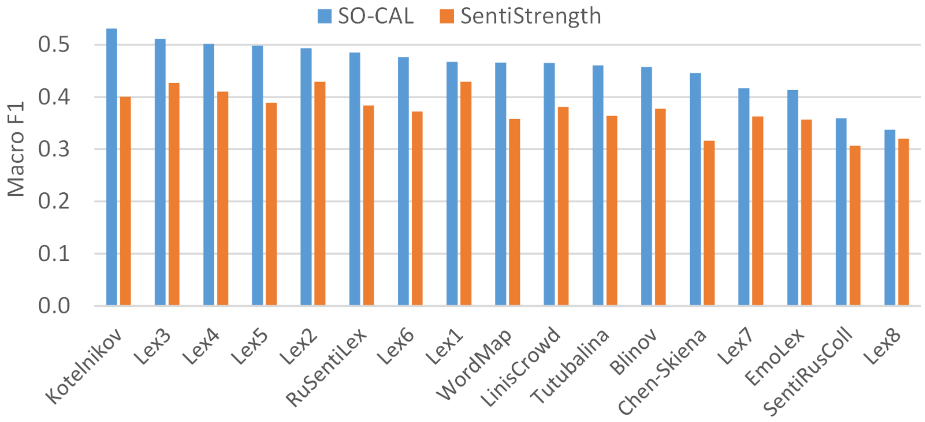


Fig. 1. The results of experiments on the selection of the optimal lexicon.

The mean values and standard deviations of the fitted thresholds for SO-CAL with Kotelnikov's lexicon turned out to be: $t_1 = -1.1 \pm 0.95$, $t_2 = 0.4 \pm 0.82$. For SentiStrength with *Lex2*, coefficient $k_{neut} = 0.6 \pm 0.65, k = 1.1 \pm 0.36$.

The results of the second series of experiments – the comparison of lexicon-based methods with RuBERT – are shown in Fig. 2. RuBERT is superior to lexicon-based methods: on average over all the corpora for RuBERT F1-score=0.5833, for SO-CAL F1 score=0.5310, for SentiStrength F1-score=0.4290.

For 12 corpora out of 16 RuBERT outperforms both lexicon-based methods. The most significant difference was for RuSentiment (28 percentage points), Romip 2012 News (21 p.p.), tweets of SentiRuEval 2015 Banks and SentiRuEval 2016 Telecoms (20 p.p.). However, for four corpora out of 16, SO-CAL comes out on top, and for three of them the difference is quite significant: SentiRuEval 2015 Cars (32 p.p.), SentiRuEval 2015 Restaurants (29 p.p.), SemEval 2016 (25 p.p.), ROMIP 2012 Books (5 p.p.). SentiStrength, as a rule, loses to both methods, with the exception of the corpus SentiRuEval 2015 Cars.

In general, RuBERT analyzes all the corpora with short texts much better – an average number of symbols in the text less than 100 (the difference is from 13 to 28 p.p.). For medium-sized texts (700-900 symbols), the lexicon-based methods are better. For longer texts, the situation is ambiguous.

5 Discussion

We have compared sets of predictions on all the test corpora for all the methods. As a result, five subsets have been obtained: 1) the predictions of all methods matched, 2) the SO-CAL and SentiStrength's predictions matched, which did not match with RuBERT; 3) the predictions from RuBERT and SO-CAL matched, which did not match with SentiStrength; 4) RuBERT and SentiStrength's predictions matched, which did not match with SO-CAL; 5) the predictions of all the methods did not match. We have calculated the macro F1-score for each of these sets (Table 4).

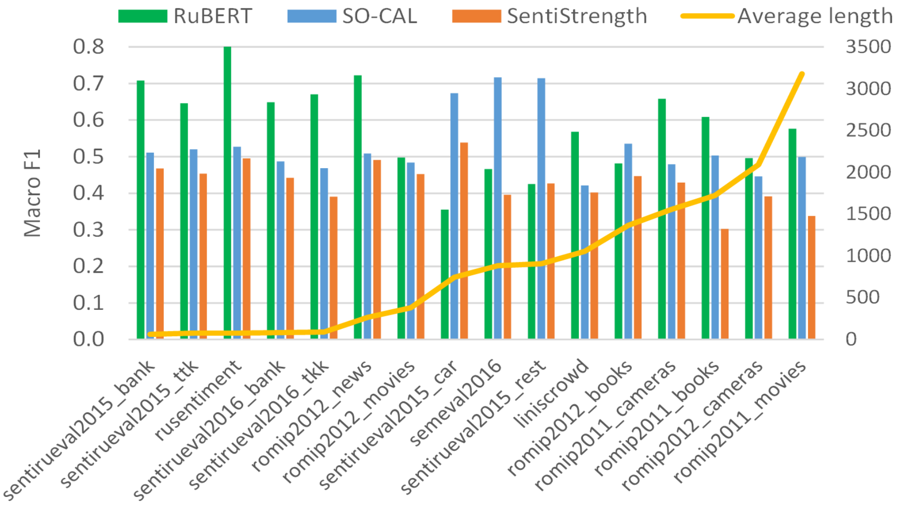


Fig. 2. Comparison results of lexicon-based methods and RuBERT on test corpora.

Table 4 shows that for the set of matching predictions (the first row – 38% of the test dataset), the performance turns out to be quite high (0.8100). On average, these are shorter texts (the average length is 481 characters). For the set of matching predictions from RuBERT and SO-CAL (the third row – 18% of the test dataset), the result is significantly higher than that of SentiStrength, which did not coincide with them – 0.7151 vs. 0.1955. For the set of identical predictions from RuBERT and SentiStrength (the fourth row – 17% of the test sample), the difference is smaller – 0.6576 vs. 0.2347 – SentiStrength is worse than SO-CAL. Finally, for the set of the unmatched predictions (the fifth row – 6%) RuBERT is far superior to both lexicon-based methods – 0.5617 vs. 0.1860 (SentiStrength) and 0.2074 (SO-CAL).

We analyzed in more detail the set of matching predictions of SO-CAL and SentiStrength that did not match RuBERT (the second row – 21% of the test dataset). The results of lexicon-based methods were significantly lower than RuBERT (0.3237 vs. 0.5625). In this case, lexicon-based methods recognize positive and negative texts poorly (0.2008 and 0.3395, respectively, versus 0.5228 and 0.6654 for RuBERT). For neutral texts, the difference is not so significant – 0.4308 for lexicon-based methods vs. 0.4993 for RuBERT. If we consider the results for individual corpora, then, in general, the picture remains the same as on Fig. 2 with a few exceptions. The lexicon-based methods show the best results for six corpora, and not for four – two new corpora (ROMIP 2012 Cameras and Movies) were added to the previous corpora (SentiRuEval 2015 Cars and Restaurants, SemEval 2016, and ROMIP 2012 Books). Also, the ROMIP 2012 Books corpus is recognized on this set of predictions much better with lexicon-based methods than with RuBERT: 0.5000 vs. 0.0833.

Table 4. Performance metrics (macro F1-score) on predictions sets.

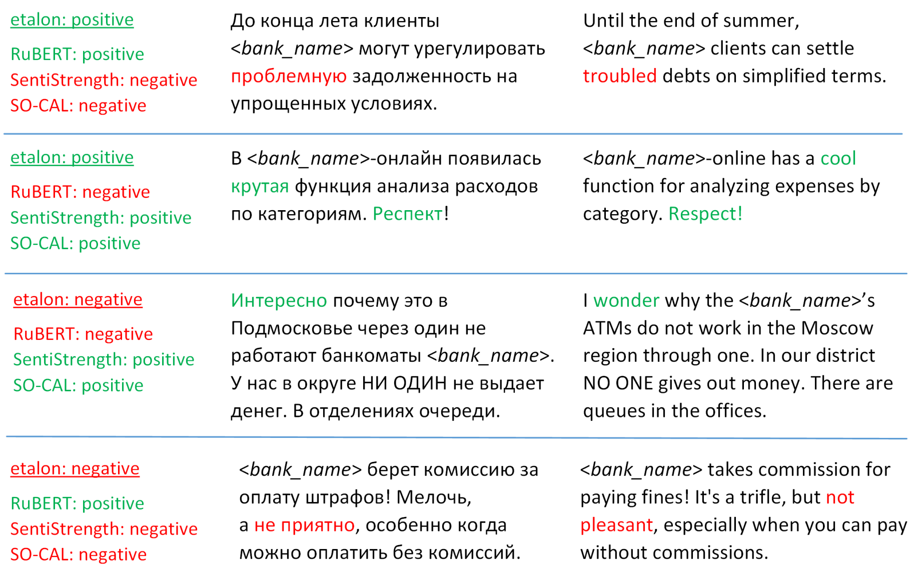
Set	RuBERT	SentiStrength	SO-CAL	Set size	Average text length, sym.
All matched	0.8100			14,310 (38%)	481
SentiStrength & SO-CAL matched	0.5625	0.3237		7,698 (21%)	519
RuBERT & SO-CAL matched	0.7151	0.1955	0.7151	6,755 (18%)	603
RuBERT & SentiStrength matched	0.6576		0.2347	6,289 (17%)	686
All didn't match	0.5617	0.1860	0.2074	2,362 (6%)	602

We also analyzed the reasons for the incorrect predictions of the lexicon-based methods. Errors most often arise due to an insufficient size of sentiment lexicon, the absence of sentiment words in the text, an incorrect recognition of negation and irrealis, an overbalance of words of the opposite sentiment, sarcasm, and erroneous identification of domain-oriented words.

As an illustration (see Fig. 3), we can give some examples that are incorrectly classified by lexicon-based methods (the first and the third examples) or RuBERT (the second and the fourth examples). In the first example lexicon-based methods didn't recognize that the phrase *settle trouble debts* has positive polarity. In the third example the word *wonder* led the lexicon-based methods to the wrong decision: they didn't take into account the phrase *ATMs do not work*. Unfortunately, RuBERT does not have the same good interpretability as lexicon-based methods, so we can't explain why RuBERT misclassified the second and fourth examples. But as we can see from the first and third examples, RuBERT can correctly classify texts even when they have words of opposite sentiment.

6 Conclusion

We have compared the lexicon-based methods SO-CAL and SentiStrength with a deep neural network model RuBERT on 16 Russian-language sentiment corpora for a three-class problem of sentiment analysis. On average, RuBERT shows



etalon: positive RuBERT: positive SentiStrength: negative SO-CAL: negative	До конца лета клиенты <*bank_name*> могут урегулировать проблемную задолженность на упрощенных условиях.	Until the end of summer, <*bank_name*> clients can settle troubled debts on simplified terms.
etalon: positive RuBERT: negative SentiStrength: positive SO-CAL: positive	В <*bank_name*>-онлайн появилась крутая функция анализа расходов по категориям. Респект!	<*bank_name*>-online has a cool function for analyzing expenses by category. Respect!
etalon: negative RuBERT: negative SentiStrength: positive SO-CAL: positive	Интересно почему это в Подмосковье через один не работают банкоматы <*bank_name*>. У нас в округе НИ ОДИН не выдает денег. В отделениях очереди.	I wonder why the <*bank_name*>'s ATMs do not work in the Moscow region through one. In our district NO ONE gives out money. There are queues in the offices.
etalon: negative RuBERT: positive SentiStrength: negative SO-CAL: negative	<*bank_name*> берет комиссию за оплату штрафов! Мелочь, а не приятно, особенно когда можно оплатить без комиссий.	<*bank_name*> takes commission for paying fines! It's a trifle, but not pleasant, especially when you can pay without commissions.

Fig. 3. Examples of classifiers' errors: the first and third examples are incorrectly classified by lexicon-based methods; the second and fourth examples are misclassified by RuBERT. The first column shows real class and the answers of the methods. The second column is in Russian, the third one – the translation into English. Sentiment words in examples are coloured: positive ones are green, negative ones are red.

a higher classification performance than lexicon-based methods, exceeding SO-CAL by an average of 5 p.p. SentiStrength lags behind SO-CAL by 10 p.p. However, for four corpora out of 16 (usually with medium-length texts) SO-CAL shows better results than RuBERT. This keeps us optimistic about the lexicon-based approach in general.

In the future, we intend to modify SO-CAL in order to more accurately take into account the peculiarities of the Russian language, such as negation and irrealis. Also, a promising area of research is the development of hybrid models that combine the ability to take into account the context of deep neural networks and linguistic knowledge contained in the sentiment lexicons.

Acknowledgement

This research is financially supported by The Russian Science Foundation, Agreement №17-71-30029 with co-financing of Bank Saint Petersburg.

References

1. Belinkov, Y., Gehrmann, S., Pavlick, E.: Interpretability and Analysis in Neural NLP. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 1–5 (2020)
2. Birjali, M., Kasri, M., Beni-Hssane, A.: A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems **226**, 107134 (2021)
3. Blinov, P.D., Klekovkina, M.V., Kotelnikov, E.V., Pestov, O.A.: Research of lexical approach and machine learning methods for sentiment analysis. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue”, vol. 12(19), pp. 51–61 (2013)
4. Chen, Y., Skiena, S.: Building Sentiment Lexicons for All Major Languages. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 383–389 (2014)
5. Chetviorkin, I., Braslavskiy, P., Loukachevitch, N.: Sentiment Analysis Track at ROMIP 2011. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog”, vol. 2, pp. 1–14 (2012)
6. Chetviorkin, I.I., Loukachevitch, N.V.: Sentiment Analysis Track at ROMIP 2012. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog”, vol. 2, pp. 40–50 (2013)
7. De Smedt, T., Daelemans, W.: Pattern for Python. Journal of Machine Learning Research **13**, 2063–2067 (2012)
8. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of 7th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), pp. 4171–4186 (2019)
9. Golubev, A., Loukachevitch, N.: Transfer Learning for Improving results on Russian Sentiment Datasets. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog”, pp. 268–277 (2021)

10. Hutto, C.J., Gilbert, E.: VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the International AAAI Conference on Web and Social Media, pp. 216–225 (2014)
11. Koltsova, O.Y., Alexeeva, S.V., Kolcov, S.N.: An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog", pp. 277–287 (2016)
12. Kotelnikov, E., Bushmeleva, N., Razova, E., Peskisheva, T., Pletneva, M.: Manually Created Sentiment Lexicons: Research and Development. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog", vol. 15(22), pp. 300–314 (2016)
13. Kotelnikov, E.V., Peskisheva, T.A., Kotelnikova, A.V., Razova, E.V.: A comparative study of publicly available russian sentiment lexicons. In: Proceedings of the 7th conference on Artificial Intelligence and Natural Language (AINL), pp. 139–151 (2018)
14. Kotelnikova, A., Kotelnikov, E.: SentiRusColl: Russian Collocation Lexicon for Sentiment Analysis. In: Artificial Intelligence and Natural Language (AINL). Communications in Computer and Information Science, vol. 1119, pp. 18–32 (2019)
15. Kulagin, D.: Russian Word Sentiment Polarity Dictionary: a Publicly Available Dataset. In: Artificial Intelligence and Natural Language. AINL 2019 (2019)
16. Kuratov, Y., Arkhipov, M.: Adaptation of deep bidirectional multilingual transformers for russian language. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog", pp. 333–340 (2019)
17. Kuznetsova, E.S., Chetviorkin, I.I., Loukachevitch, N.V.: Testing rules for sentiment analysis system. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog", vol. 2, pp. 71–80 (2013)
18. Li, H.: Deep learning for natural language processing: advantages and challenges. National Science Review **5**(1), 24–26 (2018)
19. Loukachevitch, N.V., Rubtsova, Y.V.: SentiRuEval-2016: Overcoming Time Gap and Data Sparsity in Tweet Sentiment Analysis. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog", pp. 416–426 (2016)
20. Loukachevitch, N., Levchik, A.: Creating a General Russian Sentiment Lexicon. In: Proceedings of Language Resources and Evaluation Conference (LREC), pp. 1171–1176 (2016)
21. Loukashevitch, N.V., Blinov, P.D., Kotelnikov, E.V., Rubtsova, Y.V., Ivanov, V.V., Tutubalina, E.V.: SentiRuEval: Testing Object-Oriented Sentiment Analysis Systems in Russian. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog", vol. 2, pp. 2–13 (2015)
22. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word-emotion association lexicon. Computational Intelligence **29**(3), 436–465 (2013)
23. Pontiki, M., et al.: SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval), pp. 19–30 (2016)
24. Rogers, A., Romanov, A., Rumshisky, A., Volkova, S., Gronas, M., Gribov, A.: RuSentiment: An enriched sentiment analysis dataset for social media in Russian. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 755–763 (2018)
25. Schmidt, T., Dangel, J., Wolff, C.: SentText: A Tool for Lexicon-Based Sentiment Analysis in Digital Humanities. In: Proceedings of the 16th International Symposium of Information Science (ISI), pp. 156–172 (2021)

26. Smetanin, S.: The applications of sentiment analysis for russian language texts: Current challenges and future perspectives. IEEE Access **8**, 110693–110719 (2020)
27. Smetanin, S., Komarov, M.: Deep transfer learning baselines for sentiment analysis in russian. Information Processing and Management **58**, 102484 (2021)
28. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A., Potts, C.: Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1631–1642 (2013)
29. Sun, Z., Fan, C., Han, Q., Sun, X., Meng, Y., et al.: Self-explaining Structures Improve NLP Models (2020), https://arxiv.org/abs/2012.01786
30. Taboada, M.: Sentiment Analysis: An Overview from Linguistics. Annual Review of Linguistics **2**, 325–347 (2016)
31. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics **37**(2), 267–307 (2011)
32. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment Strength Detection in Short Informal text. Journal of the American Society for Information Science and Technology **61**(12), 2544–2558 (2010)
33. Tutubalina, E.V.: Extraction and summarization methods for critical user reviews of a product. Ph.D. thesis, Kazan Federal University, Kazan, Russia (2016)
34. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is All you Need. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008 (2017)
35. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R., Le, Q.V.: XLNet: Generalized Autoregressive Pretraining for Language Understanding. In: Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), vol. 32 (2019)
36. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: Proceedings of the 29th Conference on Neural Information Processing Systems (NeurIPS), vol. 28 (2015)