

Article

Not peer-reviewed version

Leveraging Sentiment Lexicon in Sentiment Detection

Alex Johnson , Emily Davis , [Wyne Nasir](#) , Michael Brown *

Posted Date: 3 April 2024

doi: 10.20944/preprints202404.0254.v1

Keywords: Sentiment Detection; Lexicon; Enhanced SentiStrength



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Leveraging Sentiment Lexicon in Sentiment Detection

Alex Johnson, Emily Davis, Wyne Nasir and Michael Brown *

Tufts University

* Correspondence: michaelbrown@tufts.edu

Abstract: In the rapidly evolving field of sentiment analysis, the introduction of Transformer-based architectures, particularly the BERT model, has markedly improved accuracy levels, setting new standards in the analysis of textual sentiment. These advancements have been instrumental in enhancing the model's ability to grasp the nuances and complexities inherent in human language, thereby providing deeper insights into the sentiment expressed in various texts. However, the impressive performance of such deep learning models comes at the cost of increased computational demands and a lack of transparency in their decision-making processes. These challenges have reignited interest in rule-based sentiment analysis methods, which utilize sentiment lexicons for a more straightforward and computationally economical approach to determining text sentiment. Despite being overshadowed by the rise of machine learning models in recent years, these lexicon-based methods possess distinct advantages, including ease of interpretation and lower resource requirements, making them particularly appealing for certain applications. This paper seeks to re-evaluate the relevance and effectiveness of two prominent lexicon-based sentiment analysis methods, SO-CAL and SentiStrength, which have been specifically adapted for the language, in light of the advancements represented by Transformer-based models like SentBERT. Through a comprehensive comparative analysis, we examine the performance of these methodologies in contrast to SentBERT across an extensive collection of 16 text corpora, spanning a variety of genres and contexts. Our findings reveal a nuanced landscape of performance, where SentBERT's advanced capabilities typically afford it a significant advantage in accurately capturing sentiment. Nevertheless, in a surprising turn, the SO-CAL method exhibits exceptional performance on a substantial portion of the datasets, underscoring the continuing value and potential of lexicon-based approaches in sentiment analysis. This study not only highlights the strengths and weaknesses of both deep learning and lexicon-based methods but also opens the door for future hybrid approaches that could leverage the best of both worlds to achieve even greater accuracy and efficiency in sentiment analysis tasks.

Keywords: sentiment detection; lexicon; enhanced sentistrength

1. Introduction

Sentiment analysis, also known as opinion mining, has emerged as a pivotal research field that intersects natural language processing (NLP), text analysis, and computational linguistics. Its primary aim is to ascertain the sentiment within a body of text, categorizing it into categories such as positive, negative, or neutral. The significance of sentiment analysis has been magnified with the advent of social media platforms, e-commerce sites, and online forums, where vast amounts of opinion-rich data are generated daily. This surge of user-generated content has provided both opportunities and challenges for researchers and practitioners aiming to extract meaningful sentiment indicators from unstructured text. Early works in sentiment analysis focused on simple lexical approaches that relied on sentiment dictionaries or lexicons. These methods assign polarity scores to words within a text, summing these scores to predict the overall sentiment. However, such approaches often struggled with the nuances of language, such as sarcasm, idioms, and context-dependent meanings.

The introduction of machine learning models marked a significant evolution in sentiment analysis methodologies. Unlike lexicon-based approaches, machine learning models, particularly supervised learning algorithms, learn to predict sentiment by recognizing patterns in labeled training data. This shift enabled the development of models that could understand complex language features and context beyond the scope of predefined sentiment lexicons. Techniques ranging from support vector machines (SVMs) and decision trees to ensemble methods have been employed, showcasing improved accuracy

in various sentiment analysis tasks. More recently, deep learning models, especially those based on neural networks, have set new benchmarks for sentiment analysis performance. These models leverage large amounts of data to learn rich representations of text, capturing subtle linguistic cues and context that elude simpler models.

The advent of transformer-based models, epitomized by BERT (Bidirectional Encoder Representations from Transformers) and its variants, represents the cutting edge in sentiment analysis. These models employ a mechanism called attention, allowing them to focus on relevant parts of the text when predicting sentiment, which enables a nuanced understanding of complex and context-sensitive linguistic structures. Transformer models have demonstrated remarkable success in capturing the intricacies of human language, significantly advancing the field of sentiment analysis. Their ability to adapt to different languages and domains without substantial modification has further solidified their standing as a powerful tool for sentiment analysis. As sentiment analysis continues to evolve, the exploration of hybrid models that combine the strengths of rule-based and machine learning approaches, as well as the integration of multimodal data, including text, images, and video, represents promising avenues for future research.

The landscape of sentiment analysis has witnessed an exponential improvement, underscored by the following milestones. In the SST-5 corpus, accuracy metrics soared from 45.70, achieved by the RNTN framework [1], to 59.10, credited to the RoBERTa-large+Self-Explaining system [2,3]. The Yelp Reviews corpus saw a reduction in error rates from 37.95 (Char-level CNN [4]) to 27.05 (achieved by the XLNet architecture [5]). The F1-score for the ROMIP-2012 news corpus improved from 62.10, with the lexicon-based Polyarnik framework [6], to 72.69 [7]. This surge in accuracy is largely attributed to the evolution of Transformer-based models, particularly BERT [8,9], driven by deep learning advancements [10]. Despite their effectiveness, deep learning models come with their share of drawbacks: their training demands vast data volumes, high-end GPUs, and is both time and energy-intensive [11]. Moreover, the interpretability of these models remains a challenge [12].

On the flip side, lexicon or rule-based approaches [18] offer a stark contrast. They are swift, training-free, and boast high interpretability [19]. Nonetheless, the ascent of deep learning has overshadowed these lexicon-based methodologies. The linguistic domain has seen a handful of deep learning sentiment analysis studies in recent times [7,20,21]. Yet, a comparison between deep learning models and lexicon-based approaches remains unexplored. Our research endeavors to bridge this divide by juxtaposing the fine-tuned SentBERT neural network model [22] against the SO-CAL [21,23] and SentiStrength [24] methodologies, each customized for the language. Our examination spans 16 text corpora, each categorized into three sentiment classes.

The contributions of this paper are manifold:

- Adapting the SO-CAL and SentiStrength methodologies for the language, henceforth referred to as Enhanced SO-CAL and Enhanced SentiStrength.
- Conducting a comprehensive performance review of these lexicon-based approaches against SentBERT across an extensive collection of text corpora.
- Evaluating SO-CAL and SentiStrength against 17 sentiment lexicons, encompassing 9 accessible lexicons and an amalgamation of 8 additional lexicons.
- Providing a detailed analysis of the classification outcomes, highlighting the comparative strengths and weaknesses of lexicon-based strategies versus SentBERT.

2. Preliminary

Sentiment analysis has seen a variety of innovative tools and methodologies, primarily leveraging sentiment lexicons for nuanced text interpretation: Open-source offerings: Enhanced SO-CAL [23], Enhanced VADER [28], Pattern, and TextBlob; Proprietary solutions: Enhanced SentiStrength [24], and SentText [32].

Developed by Taboada et al., Enhanced SO-CAL (Semantic Orientation CALculator)¹ is a sophisticated method and tool designed for gauging sentiment in and Spanish texts [23]. It discerns sentiment by aggregating the weights of sentiment-bearing words (nouns, adjectives, verbs, and adverbs). Moreover, it employs a complex rule system to consider the impact of lexical markers like modifiers, negations, and irrealis markers, thus offering a nuanced sentiment analysis. Modifiers amplify or diminish the intensity of subsequent sentiment words. Negations reverse or modify the sentiment word's polarity, with Enhanced SO-CAL employing a shift mechanism and Enhanced VADER using inversion with a specific coefficient. Irrealis markers, signaling sentences whose sentiment might not be relevant, include modal verbs, conditional phrases, certain verbs (e.g., *expect*, *doubt*), punctuation like question marks, and quotations.

Hutto and Gilbert introduced Enhanced VADER (Valence Aware Dictionary for sEntiment Reasoning)², a tool and methodology for sentiment analysis in texts [28]. Constructed upon renowned dictionaries such as LIWC, ANEW, and General Inquirer, its lexicon was enriched through crowd-sourcing for sentiment intensity and expanded to include emoticons, acronyms, and slang. Enhanced VADER accounts for punctuation like exclamation marks, use of capital letters, modifiers, negations, and contrastive conjunctions, providing a comprehensive sentiment analysis framework.

Pattern³, a web mining library, facilitates sentiment analysis in and French by leveraging a lexicon of adjectives commonly found in product reviews, indicating its practical application in e-commerce and consumer feedback analysis [40]. TextBlob⁴ emerges as a versatile text processing library, incorporating a naive Bayesian classifier and an implementation from the Pattern library for sentiment analysis. It stands out for its simplicity and adaptability in processing textual data. Thelwall et al. crafted Enhanced SentiStrength⁵, specializing in the sentiment analysis of concise social media texts. This tool uniquely assigns both a negative and a positive score, reflecting the dual nature of sentiment expressions. Its algorithm is rooted in a sentiment lexicon with assigned intensity weights and incorporates linguistic nuances such as modifiers, negations, interrogative forms, slang, idioms, and emoticons, enriching the sentiment analysis process [24].

Schmidt et al. developed SentText⁶, a digital humanities-oriented web-based sentiment analysis tool. Initially designed for the German language and utilizing the SentiWS and BAWL-R dictionaries, SentText offers features like negation handling and result visualization, making it a valuable tool for academic research and digital humanities projects [32]. Its capability to highlight sentiment-bearing words, indicate word and overall text polarity, and facilitate sentiment comparison across texts underscores its utility in textual analysis. Among these instruments, to our knowledge, only Enhanced SentiStrength has been adapted to Russian, as indicated on its official webpage. However, a comprehensive description of these adaptations remains elusive.

In this study, we have undertaken the adaptation of Enhanced SO-CAL and Enhanced SentiStrength to the Russian context. Enhanced SO-CAL stands out among open-source sentiment analysis tools for its sophisticated approach. Despite its proprietary nature, Enhanced SentiStrength offers a straightforward adaptation process for new languages, necessitating a sentiment lexicon in the target language alongside additional linguistic resources such as lists of modifiers, negations, interrogative words, slang, and idioms, further underscoring the versatility and applicability of these tools in sentiment analysis.

¹ <https://github.com/sfu-discourse-lab/SO-CAL>.

² <https://github.com/cjhutto/vaderSentiment>.

³ <https://github.com/clips/pattern>.

⁴ <https://textblob.readthedocs.io>.

⁵ <http://sentistrength.wlv.ac.uk>.

⁶ <https://thomasschmidtur.pythonanywhere.com>.

3. Methodology

3.1. Adaptation Procedure for EmoAnalytica

The adaptation of EmoAnalytica's core methodologies, previously known as SO-CAL and SentiStrength, to accommodate the Russian linguistic nuances entails several meticulously planned steps:

- Conducting a comprehensive morphological analysis of the input texts utilizing RNNMorph, an advanced neural network model designed for high-accuracy linguistic analysis⁷;
- Compiling an extensive Russian-language sentiment lexicon, drawing upon existing lexicons while ensuring a broad representation of emotional expressions as delineated in Section 3.2. EmoAnalytica [46,47] focuses on a wide spectrum of word classes, including nouns, adjectives, verbs, and adverbs, to capture the sentiment;
- Developing detailed lists of Russian modifiers (e.g., *very*, *barely*, *significantly*) and negations (e.g., *not*, *without*, *impossible*), derived from translating and expanding existing lists from SO-CAL and SentiStrength with culturally and linguistically specific synonyms [49,50];
- Creating for EmoAnalytica, a unique set of Russian-language irrealis markers (e.g., *expect*, *might*, *anyone*), enhancing its ability to discern conditional or hypothetical sentiments;
- Refining EmoAnalytica's processing algorithms to seamlessly integrate with the results of Russian morphological analysis [49], enabling a deeper and more nuanced understanding of sentiment in Russian texts;
- Establishing an efficient programming interface with EmoAnalytica's desktop application to facilitate the submission and analysis of input texts, streamlining the sentiment analysis process [49].

3.2. Sentiment Lexicons in Detail

Central to EmoAnalytica's analytical prowess is its repository of sentiment lexicons, the depth, and breadth of which significantly influence the accuracy and reliability of sentiment analysis outcomes. We have meticulously curated 9 publicly accessible Russian sentiment lexicons, including EmoLex and Chen-Skiena's, which offer Russian adaptations of globally recognized multi-lingual lexicons [56].

To enhance the quality and consistency of these lexicons, we undertook the following normalization steps:

- Exclusion of neutral words to sharpen the focus on distinctly positive or negative expressions;
- Removal of ambiguous words that could be interpreted as both positive and negative, ensuring clarity in sentiment analysis;
- Elimination of words incorporating Latin characters to maintain linguistic purity and relevance;
- Standardization of all entries to lowercase to avoid duplications or inconsistencies caused by case variations;
- Application of RNNMorph for word normalization, ensuring uniformity and accuracy in lexical entries;
- Ensuring uniqueness within the lexicon by eliminating duplicate entries, whether they are individual words or phrases, to maintain a clean and efficient analytical database.

The compilation and refinement of these lexicons were undertaken with meticulous care to ensure their comprehensive coverage and applicability in sentiment analysis, as summarized in Table 1.

⁷ <https://github.com/IlyaGusev/rnnmorph>.

Table 1. Comprehensive Characteristics of Russian Sentiment Lexicons Utilized in EmoAnalytica.

Lexicon	Total	Positive Elements		Negative Elements	
RuSentiLex	12,560	3,258	25.9%	9,302	74.1%
Word Map	11,237	4,491	40.0%	6,746	60.0%
SentiRusColl	6,538	3,981	60.9%	2,557	39.1%
EmoLex	4,600	1,982	43.1%	2,618	56.9%
LinisCrowd	3,986	1,126	28.2%	2,860	71.8%
Blinov’s Lexicon	3,524	1,611	45.7%	1,913	54.3%
Kotelnikov’s Lexicon	3,206	1,028	32.1%	2,178	67.9%
Chen-Skienna’s Lexicon	2,604	1,139	43.7%	1,465	56.3%
Tutubalina’s Lexicon	2,442	1,032	42.3%	1,410	57.7%

To further enhance EmoAnalytica’s analytical capabilities, we employed a "voting" mechanism across these lexicons to create a series of 8 combined lexicons, *Lex1* through *Lex8*, each incorporating words featured in at least *N* sentiment lexicons. This innovative approach allowed us to distill a highly refined and consensus-based sentiment lexicon, with *Lex1* encompassing the broadest array of sentiment expressions and *Lex9* being notably empty, underscoring the absence of unanimous agreement across all lexicons. The attributes of these combined lexicons are meticulously cataloged in Table 2.

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%

To confront and rectify the class imbalance inherent in emotion detection datasets, we have devised a customized version of the Binary Cross-Entropy (BCE) Loss, aptly suited for the nuances of multi-label classification. This refined loss function is expressed as:

$$L = \sum_{n=1}^N \sum_{c=1}^C -w_c [y_{n,c} \log(\sigma(x_{n,c})) + (1 - y_{n,c}) \log(1 - \sigma(x_{n,c}))]$$

(1)

In this formula, *N* symbolizes the batch size, *C* represents the total number of classes, *x_{n,c}* indicates the output of the model for class *c* of sample *n*, and *w_c* signifies the class-specific positive weighting factor, determined as:

$$w_c = \frac{\text{number of negative samples for class } c}{\text{number of positive samples for class } c}$$

This weighting mechanism, derived from the empirical data of the training set, is ingeniously calibrated to enhance recall in instances dominated by negative samples for class *c*, and bolster precision in scenarios where positive samples prevail.

Although explorations into adapting the focal loss for multi-label classification were conducted, these investigations concluded that it did not offer a significant advantage over the traditional BCE loss in terms of augmenting our model’s efficacy. Through this detailed exposition, EmoLeverage stands unveiled as a beacon of innovation in the NLP landscape, charting a new course for emotion detection with its ground-breaking architecture and meticulously crafted methodologies. Our ambition with

EmoLeverage is not only to surpass existing benchmarks but to redefine them, establishing a new standard of excellence in the detection and analysis of emotional expressions in text.

4. Experiments

4.1. Configurations

In this study, we evaluated the performance of two sentiment analysis methods tailored for the Russian language: LexiRus and SentiNet, alongside a deep learning model named SentLSTM [22]. Traditionally, lexicon-based techniques offer a viable option for sentiment analysis tasks without necessitating extensive training. However, leveraging available training corpora, we opted to fine-tune the hyperparameters of these lexicon-based methods. Specifically, we meticulously selected the most suitable sentiment lexicons (out of 17 lexicons – see Section 3.2) and determined the thresholds for classifying sentiments.

The thresholds are computed as follows. LexiRus computes a singular sentiment score s for a given text, which is then categorized into one of three classes (positive, negative, or neutral). We derived two thresholds, t_{pos} and t_{neg} , from the training data. The sentiment classification for a text c is determined based on the following conditions:

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

On the other hand, SentiNet yields two sentiment scores for a text – positive s_{pos} and negative s_{neg} . We selected coefficients k_{neut} and k such that:

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

Further details regarding lexicon selection and threshold determination are presented in Section ??.

The pretrained SentLSTM model underwent fine-tuning individually on each training corpus with the subsequent configuration: learning rate $2 \cdot 10^{-5}$, 5 epochs, and a batch size of 12. The reported results are averages over five runs to mitigate the impact of random weight initialization. Training was conducted leveraging the Google Colab Pro platform, utilizing NVIDIA Tesla P100 and V100 GPUs. Across all corpora, sentiment analysis was approached as a three-class problem, involving the classification of texts into positive, negative, and neutral categories. The primary evaluation metric employed was the macro F1-score.

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%

Table 3. Cont.

Corpus	Type	Split	Total	Positive	Negative	Neutral
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%
SentiRuEval 2015	car reviews	test	4,573	31.7%	41.3%	27.0%
		train	203	56.6%	14.8%	28.6%
	restaurant reviews	test	200	49.0%	13.0%	38.0%
		train	200	68.0%	14.0%	18.0%
	bank tweets	test	203	71.9%	12.8%	15.3%
		train	4,883	7.2%	21.7%	71.1%
	telecom tweets	test	4,534	7.6%	14.4%	78.0%
		train	4,839	18.8%	32.7%	48.5%
SentiRuEval 2016	bank tweets	test	3,774	9.1%	22.4%	68.5%
	telecom tweets	test	3,302	9.1%	23.1%	67.8%
SemEval 2016 RuSentiment	restaurant reviews	test	2,198	8.3%	45.8%	45.9%
	posts	test	103	67.0%	14.6%	18.4%
		train	24,124	38.0%	15.2%	46.8%
		test	2,621	36.0%	9.8%	54.2%

4.2. Experimental Results

Two sets of experiments were conducted. Initially, the training data were utilized to optimize the hyperparameters for the lexicon-based methods, including lexicon selection and threshold determination (see Section ??). Subsequently, in the second series of experiments, the lexicon-based methods, with optimized hyperparameters, were compared against the fine-tuned SentLSTM model using test data.

The outcomes from the first series of experiments concerning the optimal sentiment lexicon selection are depicted in Figure ?. Kotelnikov's lexicon emerged as the most effective choice for LexiRus, while *Lex1* and *Lex2* were identified as optimal for SentiNet (with *Lex2* being favored due to its compact size).

Mean values and standard deviations for the derived thresholds were as follows: For LexiRus with Kotelnikov's lexicon, $t_1 = -1.1 \pm 0.95$ and $t_2 = 0.4 \pm 0.82$. For SentiNet with *Lex2*, coefficients were $k_{neut} = 0.6 \pm 0.65$ and $k = 1.1 \pm 0.36$.

The second series of experiments, comparing lexicon-based methods with SentLSTM. SentLSTM exhibited superior performance over the lexicon-based methods, with an average F1-score of 0.5833 compared to 0.5310 for LexiRus and 0.4290 for SentiNet.

SentLSTM outperformed both lexicon-based methods for 12 out of 16 corpora. The disparities were most notable in RuSentiment (28 percentage points), Romip 2012 News (21 p.p.), and tweets from SentiRuEval 2015 Banks and SentiRuEval 2016 Telecoms (20 p.p.). However, in four corpora, LexiRus demonstrated superior performance, with significant differences observed in SentiRuEval 2015 Cars (32 p.p.), SentiRuEval 2015 Restaurants (29 p.p.), SemEval 2016 (25 p.p.), and ROMIP 2012 Books (5 p.p.). SentiNet consistently lagged behind both methods, except for the SentiRuEval 2015 Cars corpus.

Overall, SentLSTM excelled in analyzing corpora with shorter texts (averaging fewer than 100 symbols), exhibiting differences ranging from 13 to 28 percentage points. For texts of medium length (700-900 symbols), lexicon-based methods displayed superior performance. However, for longer texts, the comparative efficacy was less definitive.

4.3. Analysis

In our analysis, we juxtaposed diverse prediction sets across all test corpora for each method. This process yielded five distinct subsets: 1) predictions unanimously aligned across all methods, 2) predictions congruent between SO-CAL and SentiStrength but not with SentBERT, 3) predictions matching between SentBERT and SO-CAL but not with SentiStrength, 4) predictions matching between

SentBERT and SentiStrength but not with SO-CAL, and 5) predictions diverging across all methods. We computed the macro F1-score for each subset (Table 4).

Table 4 illustrates that the subset with aligned predictions (first row – comprising 38% of the test dataset) achieved commendable performance (0.8100). Notably, these predictions predominantly pertain to shorter texts, averaging 481 characters. For the subset where predictions matched between SentBERT and SO-CAL (third row – constituting 18% of the test dataset), the performance substantially outstripped that of SentiStrength, which failed to coincide with them – 0.7151 vs. 0.1955. In the subset with congruent predictions from SentBERT and SentiStrength (fourth row – encompassing 17% of the test sample), the performance gap was narrower – 0.6576 vs. 0.2347 – with SentiStrength trailing SO-CAL. Conversely, in the subset of unmatched predictions (fifth row – representing 6% of the dataset), SentBERT significantly outperformed both lexicon-based methods – 0.5617 vs. 0.1860 (SentiStrength) and 0.2074 (SO-CAL).

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

Case	SentBERT	Senti-Strength	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
SentBERT & SO-CAL matched	0.7151	0.1955	0.7151	6,755 (18%)	603
SentBERT & Senti-Strength matched	0.6576		0.2347	6,289 (17%)	686
All didn't match	0.5617	0.1860	0.2074	2,362 (6%)	602

We conducted a detailed analysis of the subset where predictions from SO-CAL and SentiStrength matched but diverged from SentBERT (second row – constituting 21% of the test dataset). Results from the lexicon-based methods were markedly inferior to SentBERT (0.3237 vs. 0.5625). Particularly, lexicon-based methods exhibited poor recognition of positive and negative texts (0.2008 and 0.3395, respectively, compared to 0.5228 and 0.6654 for SentBERT). Although the disparity for neutral texts was less pronounced – 0.4308 for lexicon-based methods vs. 0.4993 for SentBERT. Examining results for individual corpora revealed a consistent pattern, largely aligning with the trends, with a few exceptions. Lexicon-based methods demonstrated optimal performance for six corpora, excluding four – two new corpora (ROMIP 2012 Cameras and Movies) alongside the previous set (SentiRuEval 2015 Cars and Restaurants, SemEval 2016, and ROMIP 2012 Books). Remarkably, the ROMIP 2012 Books corpus exhibited significantly better recognition with lexicon-based methods compared to SentBERT on this subset of predictions: 0.5000 vs. 0.0833.

Additionally, we delved into the causes behind erroneous predictions made by the lexicon-based methods. Errors predominantly stem from factors such as limited sentiment lexicon size, absence of sentiment-laden words in the text, misinterpretation of negation and irrealis, disproportionate presence of opposing sentiment words, sarcasm, and erroneous identification of domain-specific terms. To elucidate further, we present examples misclassified by either lexicon-based methods (first and third examples) or SentBERT (second and fourth examples). In the first instance, lexicon-based methods failed to discern the positive polarity within the phrase *settle trouble debts*. Conversely, in the third example, the inclusion of the word *wonder* led to an erroneous decision, as lexicon-based methods overlooked the negation in *ATMs do not work*. Unfortunately, SentBERT's interpretability doesn't offer comparable insights, hindering the explanation of misclassifications in the second and fourth examples. Nonetheless, as evidenced by the first and third instances, SentBERT demonstrates adeptness in correctly classifying texts even when featuring words with conflicting sentiments.

5. Conclusions and Future Work

The comparative analysis conducted between the lexicon-based methodologies SO-CAL and SentiStrength against the deep neural architecture SentBERT across 16 Russian-language sentiment corpora for sentiment analysis revealed intriguing insights. In the pursuit of categorizing sentiments into three distinct classes, SentBERT consistently demonstrated superior classification performance, surpassing SO-CAL by an average of 5 percentage points, while SentiStrength lagged behind SO-CAL by 10 percentage points. Nevertheless, SO-CAL exhibited superior performance over SentBERT in four out of 16 corpora, particularly those featuring texts of medium length. This observation instills confidence in the viability of the lexicon-based approach within the broader landscape of sentiment analysis.

Looking ahead, future endeavors will focus on refining the SO-CAL framework to incorporate a more nuanced understanding of the intricacies inherent in the Russian language, such as negation and irrealis. Additionally, the exploration of hybrid models emerges as a promising avenue for research, aimed at synergizing the contextual awareness of deep neural networks with the linguistic expertise encapsulated within sentiment lexicons. The future trajectory of our research endeavors encompasses several key avenues for exploration and enhancement:

- **Enhanced Linguistic Considerations:** Further refinement of SO-CAL to intricately account for linguistic nuances specific to the Russian language, including nuanced expressions of negation and irrealis, aiming to bolster its performance across diverse linguistic contexts.
- **Hybrid Model Development:** Pursuing the development of hybrid models that seamlessly integrate deep neural network architectures with the comprehensive linguistic knowledge embedded within sentiment lexicons. This hybridization seeks to leverage the strengths of both paradigms to achieve enhanced sentiment analysis accuracy and robustness.
- **Domain-Specific Adaptation:** Investigating strategies for domain-specific adaptation of sentiment analysis models, tailoring their capabilities to effectively address the unique linguistic characteristics and sentiment dynamics prevalent in specialized domains such as finance, health-care, social media, and customer reviews.
- **Multimodal Sentiment Analysis:** Venturing into the realm of multimodal sentiment analysis by integrating textual, visual, and auditory modalities, thereby enriching the analysis process with a comprehensive understanding of sentiments conveyed through diverse channels of communication. This multidimensional approach promises to capture a more holistic view of sentiment, accommodating the inherent complexities and nuances present in multimodal data sources.
- **Explainable AI for Sentiment Analysis:** Exploring the integration of explainable AI techniques within sentiment analysis models to enhance transparency and interpretability. By providing insights into the decision-making process of sentiment analysis algorithms, explainable AI facilitates user trust and comprehension, enabling stakeholders to make informed decisions based on sentiment analysis outcomes.
- **Continual Learning Frameworks:** Investigating continual learning frameworks for sentiment analysis models to adapt and evolve over time in response to changing linguistic patterns, domain-specific shifts, and evolving user preferences. Continual learning enables sentiment analysis systems to remain relevant and effective in dynamic environments, ensuring sustained performance and adaptability.
- **Ethical Considerations in Sentiment Analysis:** Addressing ethical considerations and biases inherent in sentiment analysis algorithms, particularly concerning issues of fairness, transparency, and accountability. By prioritizing ethical guidelines and frameworks, we aim to develop sentiment analysis models that uphold principles of equity, justice, and societal responsibility.
- **Cross-Lingual Sentiment Analysis:** Exploring methodologies for cross-lingual sentiment analysis to extend sentiment analysis capabilities across diverse linguistic contexts and languages. Cross-lingual sentiment analysis facilitates the analysis of sentiments expressed in multiple languages, enabling broader insights into global sentiment trends and cross-cultural dynamics.

Through the pursuit of these avenues, we anticipate fostering advancements in the field of sentiment analysis, culminating in the development of more robust, adaptable, and linguistically aware models capable of effectively capturing and interpreting sentiments across diverse linguistic and contextual landscapes.

References

1. 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. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013, pp. 1631–1642.
2. Sun, Z.; Fan, C.; Han, Q.; Sun, X.; Meng, Y.; others. Self-Explaining Structures Improve NLP Models **2020**.
3. Fei, H.; Zhang, M.; Ji, D. Cross-Lingual Semantic Role Labeling with High-Quality Translated Training Corpus. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7014–7026.
4. Zhang, X.; Zhao, J.; LeCun, Y. Character-level Convolutional Networks for Text Classification. *Proceedings of the 29th Conference on Neural Information Processing Systems (NeurIPS)*, 2015, Vol. 28.
5. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019, Vol. 32.
6. Kuznetsova, E.S.; Chetviorkin, I.I.; Loukachevitch, N.V. Testing rules for sentiment analysis system. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog”*, 2013, Vol. 2, pp. 71–80.
7. Golubev, A.; Loukachevitch, N. Transfer Learning for Improving Results on Russian Sentiment Datasets. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog”*, 2021, pp. 268–277.
8. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of 7th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, 2019, pp. 4171–4186.
9. Wu, S.; Fei, H.; Qu, L.; Ji, W.; Chua, T.S. NExT-GPT: Any-to-Any Multimodal LLM. *CoRR* **2023**, *abs/2309.05519*.
10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*, 2017, Vol. 30, pp. 5998–6008.
11. Li, H. Deep learning for natural language processing: advantages and challenges. *National Science Review* **2018**, *5*, 24–26.
12. Belinkov, Y.; Gehrmann, S.; Pavlick, E. Interpretability and Analysis in Neural NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020., pp. 1–5.
13. Wu, S.; Fei, H.; Li, F.; Zhang, M.; Liu, Y.; Teng, C.; Ji, D. Mastering the Explicit Opinion-Role Interaction: Syntax-Aided Neural Transition System for Unified Opinion Role Labeling. *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence*, 2022, pp. 11513–11521.
14. Shi, W.; Li, F.; Li, J.; Fei, H.; Ji, D. Effective Token Graph Modeling using a Novel Labeling Strategy for Structured Sentiment Analysis. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 4232–4241.
15. Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent Emotion Memory for Multi-Label Emotion Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 7692–7699.
16. Wang, F.; Li, F.; Fei, H.; Li, J.; Wu, S.; Su, F.; Shi, W.; Ji, D.; Cai, B. Entity-centered Cross-document Relation Extraction. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022, pp. 9871–9881.
17. Zhuang, L.; Fei, H.; Hu, P. Knowledge-enhanced event relation extraction via event ontology prompt. *Inf. Fusion* **2023**, *100*, 101919.
18. Taboada, M. Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics* **2016**, *2*, 325–347.

19. Birjali, M.; Kasri, M.; Beni-Hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems* **2021**, *226*, 107134.
20. Smetanin, S.; Komarov, M. Deep transfer learning baselines for sentiment analysis in Russian. *Information Processing and Management* **2021**, *58*, 102484.
21. Fei, H.; Ren, Y.; Ji, D. Retrofitting Structure-aware Transformer Language Model for End Tasks. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 2151–2161.
22. Kuratov, Y.; Arkhipov, M. Adaptation of deep bidirectional multilingual transformers for Russian language. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog”, 2019, pp. 333–340.
23. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics* **2011**, *37*, 267–307.
24. 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* **2010**, *61*, 2544–2558.
25. Fei, H.; Wu, S.; Li, J.; Li, B.; Li, F.; Qin, L.; Zhang, M.; Zhang, M.; Chua, T.S. LasUIE: Unifying Information Extraction with Latent Adaptive Structure-aware Generative Language Model. Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022, 2022, pp. 15460–15475.
26. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D.; Liang, X. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics* **2021**, *22*.
27. Wu, S.; Fei, H.; Ji, W.; Chua, T.S. Cross2StrA: Unpaired Cross-lingual Image Captioning with Cross-lingual Cross-modal Structure-pivoted Alignment. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 2593–2608.
28. Hutto, C.J.; Gilbert, E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 2014, pp. 216–225.
29. Li, J.; Xu, K.; Li, F.; Fei, H.; Ren, Y.; Ji, D. MRN: A Locally and Globally Mention-Based Reasoning Network for Document-Level Relation Extraction. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 1359–1370.
30. Fei, H.; Wu, S.; Ren, Y.; Zhang, M. Matching Structure for Dual Learning. Proceedings of the International Conference on Machine Learning, ICML, 2022, pp. 6373–6391.
31. Cao, H.; Li, J.; Su, F.; Li, F.; Fei, H.; Wu, S.; Li, B.; Zhao, L.; Ji, D. OneEE: A One-Stage Framework for Fast Overlapping and Nested Event Extraction. Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 1953–1964.
32. Schmidt, T.; Dangel, J.; Wolff, C. SentText: A Tool for Lexicon-Based Sentiment Analysis in Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI), 2021, pp. 156–172.
33. Fei, H.; Li, F.; Li, B.; Ji, D. Encoder-Decoder Based Unified Semantic Role Labeling with Label-Aware Syntax. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 12794–12802.
34. Li, B.; Fei, H.; Li, F.; Wu, Y.; Zhang, J.; Wu, S.; Li, J.; Liu, Y.; Liao, L.; Chua, T.S.; Ji, D. DiaASQ: A Benchmark of Conversational Aspect-based Sentiment Quadruple Analysis. Findings of the Association for Computational Linguistics: ACL 2023, 2023, pp. 13449–13467.
35. Fei, H.; Ren, Y.; Ji, D. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management* **2020**, *57*, 102311.
36. Li, J.; Fei, H.; Liu, J.; Wu, S.; Zhang, M.; Teng, C.; Ji, D.; Li, F. Unified Named Entity Recognition as Word-Word Relation Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 10965–10973.
37. Wu, S.; Fei, H.; Ren, Y.; Ji, D.; Li, J. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, pp. 3957–3963.
38. Li, B.; Fei, H.; Liao, L.; Zhao, Y.; Teng, C.; Chua, T.; Ji, D.; Li, F. Revisiting Disentanglement and Fusion on Modality and Context in Conversational Multimodal Emotion Recognition. Proceedings of the 31st ACM International Conference on Multimedia, MM, 2023, pp. 5923–5934.
39. Fei, H.; Liu, Q.; Zhang, M.; Zhang, M.; Chua, T.S. Scene Graph as Pivoting: Inference-time Image-free Unsupervised Multimodal Machine Translation with Visual Scene Hallucination. Proceedings of the 61st

- Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 5980–5994.
40. De Smedt, T.; Daelemans, W. Pattern for Python. *Journal of Machine Learning Research* **2012**, *13*, 2063–2067.
 41. Wu, S.; Fei, H.; Cao, Y.; Bing, L.; Chua, T.S. Information Screening whilst Exploiting! Multimodal Relation Extraction with Feature Denoising and Multimodal Topic Modeling. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 14734–14751.
 42. Fei, H.; Wu, S.; Ren, Y.; Li, F.; Ji, D. Better Combine Them Together! Integrating Syntactic Constituency and Dependency Representations for Semantic Role Labeling. Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, 2021, pp. 549–559.
 43. Wu, S.; Fei, H.; Zhang, H.; Chua, T.S. Imagine That! Abstract-to-Intricate Text-to-Image Synthesis with Scene Graph Hallucination Diffusion. *Advances in Neural Information Processing Systems* **2024**, *36*.
 44. Fei, H.; Wu, S.; Ji, W.; Zhang, H.; Chua, T.S. Empowering dynamics-aware text-to-video diffusion with large language models. *arXiv preprint arXiv:2308.13812* **2023**.
 45. Qu, L.; Wu, S.; Fei, H.; Nie, L.; Chua, T.S. Layoutlm-t2i: Eliciting layout guidance from llm for text-to-image generation. Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 643–654.
 46. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* **2018**, *abs/1810.04805*, [1810.04805].
 47. Ando, R.K.; Zhang, T. A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data. *Journal of Machine Learning Research* **2005**, *6*, 1817–1853.
 48. Fei, H.; Li, F.; Li, C.; Wu, S.; Li, J.; Ji, D. Inheriting the Wisdom of Predecessors: A Multiplex Cascade Framework for Unified Aspect-based Sentiment Analysis. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI, 2022, pp. 4096–4103.
 49. Pfeiffer, J.; Kamath, A.; Rücklé, A.; Cho, K.; Gurevych, I. AdapterFusion: Non-Destructive Task Composition for Transfer Learning. *CoRR* **2020**, *abs/2005.00247*, [2005.00247].
 50. Wang, G.; Ying, R.; Huang, J.; Leskovec, J. Improving Graph Attention Networks with Large Margin-based Constraints. NeurIPS-Workshop, 2019.
 51. Fei, H.; Chua, T.; Li, C.; Ji, D.; Zhang, M.; Ren, Y. On the Robustness of Aspect-based Sentiment Analysis: Rethinking Model, Data, and Training. *ACM Transactions on Information Systems* **2023**, *41*, 50:1–50:32.
 52. Zhao, Y.; Fei, H.; Cao, Y.; Li, B.; Zhang, M.; Wei, J.; Zhang, M.; Chua, T. Constructing Holistic Spatio-Temporal Scene Graph for Video Semantic Role Labeling. Proceedings of the 31st ACM International Conference on Multimedia, MM, 2023, pp. 5281–5291.
 53. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D. Nonautoregressive Encoder-Decoder Neural Framework for End-to-End Aspect-Based Sentiment Triplet Extraction. *IEEE Transactions on Neural Networks and Learning Systems* **2023**, *34*, 5544–5556.
 54. Zhao, Y.; Fei, H.; Ji, W.; Wei, J.; Zhang, M.; Zhang, M.; Chua, T.S. Generating Visual Spatial Description via Holistic 3D Scene Understanding. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 7960–7977.
 55. Fei, H.; Li, B.; Liu, Q.; Bing, L.; Li, F.; Chua, T.S. Reasoning Implicit Sentiment with Chain-of-Thought Prompting. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2023, pp. 1171–1182.
 56. Kotelnikov, E.V.; Peskicheva, T.A.; Kotelnikova, A.V.; Razova, E.V. A comparative study of publicly available Russian sentiment lexicons. Proceedings of the 7th conference on Artificial Intelligence and Natural Language (AINL), 2018, pp. 139–151.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.