Extracting Sentiments towards COVID-19 Aspects

Eduard Nugamanov, Natalia Loukachevitch, and Boris Dobrov

Lomonosov Moscow State University, Leninskie Gory, 1, Moscow, Russia

ed.nugamanov@gmail.com

Abstract. In this paper, we introduce a specialized Russian dataset and study approaches for aspect-based sentiment analysis of Russian users' comments about the COVID-19. We solve two tasks, namely Relevance Determination (RD), which aims to predict whether a sentence is relevant to an aspect of the pandemic, and Sentiment Classification (SC), which classifies the sentiment expressed towards an aspect in a sentence. We applied and tested various methods of machine learning, including fine-tuning of the pre-trained RuBERT model. The best results in both tasks were obtained by RuBERT model in the Natural Language Inference (NLI) formulation.

Keywords: Aspect-based sentiment analysis · BERT model · natural-language inference.

1 Introduction

COVID-19 is a dangerous infectious disease caused by the SARS-CoV-2 virus. Nowadays this infection is declared a pandemic and is one of the main threats to humanity endangering both physical and mental health of people. COVID-related issues are widely discussed in social media. Such discussions give great opportunities for psychologists, social scientists to study information dissemination in social networks, influence of various sources on forming users' opinions [2, 1].

Extracting opinions related to coronavirus can be considered as the aspect-based sentiment analysis task (ABSA)[17], which allows identifying sentiment towards specific issues of coronavirus epidemics. The ABSA task, intended for extraction of sentiment towards specific aspects of an entity or a topic was mainly studied on users' reviews such as restaurant reviews, for example food or service aspects. In fact, in coronavirus-oriented discussions we can see the same ABSA task. Aspect-based Sentiment Analysis applied to COVID-related messages is one of the means to reveal the most frustrating aspects of the pandemic.

In this paper, we introduce a Russian dataset and an approach to aspect-based sentiment analysis of Russian users' comments about the COVID-19. The dataset is large enough (about 10 thousand messages) to train modern machine learning methods in order to classify the flow of user opinions on the above and similar issues. A similar dataset could not be found in the current world publications. For the Russian language, there is no other manually annotated dataset of user messages related to the issues of coronavirus infection.

We solve two tasks, namely Relevance Determination (RD), which aims to predict whether a sentence is relevant to an aspect of the pandemic, and Sentiment Classification (SC), which classifies the sentiment expressed towards an aspect in a sentence. We applied and tested various methods of machine learning, including fine-tuning of the pre-trained Russian BERT, RuBERT [12] model. In addition to this, we formulate original tasks as Natural Language Inference (NLI) and Question Answering (QA) problems [23] and applied RuBERT to them, which led to a significant increase in the quality of classification.

2 Related Work

2.1 COVID-related Sentiment Analysis

During last year a lot of work were devoted to users' posts concerning COVID-19. In [21], the authors examine the propagation of misinformation, conducted sentiment analysis, and determined the main topics of discussion in a collection of tweets about the COVID-19 pandemic. The paper [3] studies people's reaction to lockdown in India with Twitter. The researchers from [20] use clustering and sentiment analysis to categorize tweets about masks.

Among used methods for sentiment analysis of COVID-related texts, general sentiment analysis prevails based on existing general sentiment classifiers [1, 2, 4, 5]. The most commonly used systems for sentiment analysis are VADER [9] and TextBlob [13]. However, the authors of [8] showed that the quality of classification of the users' sentiment into three classes in relation to vaccines by the general-purpose VADER system is about 0.51 accuracy, the TextBlob system result is slightly higher than about 0.53. These low results can be explained by the fact that the above-mentioned systems were built and trained without taking into account the specifics of the COVID-epidemic topic.

There are very few new specialized datasets that are manually annotated with respect to coronavirus or related aspects. The authors of [7] previously annotated a dataset of tweets about the attitudes of social media users towards influenza vaccines, and this set is currently being used to research users' attitudes towards coronavirus vaccination (FVD dataset). Hussain et al. [8] collected comments from Facebook and Twitter users regarding coronavirus vaccination and created a UKCOVID tagged dataset. They propose a combined approach using the VADER and TextBLob systems and retrain the BERT neural network model.

For the Russian language, about 3903 tweets were extracted in the work [30], a general sentiment analysis was carried out based on the Dostoevsky model [1]. It is important to note that the Dostoevsky model is trained on the RuSentiment dataset of VKontakte posts about Ukrainian-Russian relations [19] and does not concern medical topics. The work [15] examines the attitudes of physicians to the problems of the coronavirus epidemic in specialized medical forums in Russian.

[1] https://github.com/bureaucratic-labs/dostoevsky

2.2 ABSA Sentiment Analysis Task

ABSA determines the sentiment expressed to some aspect of an entity in a text. Typically, one aspect can be represented by several terms or can be not expressed in a text at all. Early approaches to the ABSA task utilize extensive feature-engineering. So, in [11] a sentiment score computed on a large unlabeled corpus of reviews is assigned to every word and used as an input to the Support Vector Machine classifier along with other textual and syntactical features.

Neural networks allowed researchers to avoid manual feature-engineering. First models were based on the LSTM architecture and attention mechanism. For instance, TD-LSTM [24] uses two LSTM networks to model left and right contexts of an aspect term. ATAE-LSTM [27] creates a representation of an aspect term to use in the attention mechanism along with other tokens.

Introduction of transformers [25] allowed to improve the results. Its basis, Multi-Head Self-Attention (MHSA) layers, became a popular choice to extract relations between tokens in texts. So, AEN [22] uses MHSA layers to model both a context and an aspect term in the context.

The latest innovation in NLP tasks is the utilization of pre-trained generative language models, such as ELMo [16] and BERT [6]. The latter is a bidirectional encoder based on the transformer architecture. It forms powerful context-aware representations of tokens, that can be used as an input to other architectures. Also, BERT can be fine-tuned by adding task-specific layers on top. For example, the LCF-ATEPC model [28] uses MHSA blocks on top of the BERT encodings to extract and classify target terms simultaneously. The SDGCN [29] architecture uses BERT representations as an input to BiLSTM network with Attention Mechanism, which models relations between a sentence and each its target with the help of graph convolutional networks, which model relations between different targets.

One of the most important problems that face researchers is the lack of labeled data. There are different approaches to that problem. For instance, BERT-ADA [18] performs domain adaption by pre-training BERT on unlabeled data. The BAT model [10] generates additional adversarial examples while training. The Snippext system [14] uses BERT for a variety of tasks: extraction and verification of pairs (target, opinion on target) from a text, its sentiment classification, determination of the aspect of the target. The authors utilized such techniques as data augmentation and semi-supervised learning. Besides, to perform an effective and reliable augmentation they adapted the MixUp [26] operation from computer vision.

3 Covid Aspect Sentiment Dataset

3.1 Dataset Annotation

For the dataset, users' comments on Covid-2019 related news articles were collected from the VKontakte social network. We selected masks, quarantine (lockdown) or vaccines as aspects for sentiment annotation and extracted relevant comments using corresponding keywords. Also sentiment attitudes towards government measures were annotated for all selected comments. This government aspect is especially difficult for automatic analysis because mentioning of government can be implicit as in the following sentence:

- *In Germany, a permanent mask, etc. regime, shots from Russia are very surprising, when nothing is observed at all.*

The total number of sentences is 10968.

Each sentence was labeled by several experts (three on average). An annotator should indicate sentiment it expresses towards each of the above-mentioned four aspects (or indicate that the sentence is not relevant to the aspect). The annotators' group included professional linguists and psychologists. We consider six types of sentiment labels, namely:

- irrelevant;
- positive;
- negative;
- neutral. This label is used for factual sentences without any visible sentiments;
- both positive and negative. For such a label, evident positive and negative attitudes should be seen in a message;
- relevant, but impossible to determine. In this case, the presence of a sentiment attitude is seen, but the context of sentence does not give possibility to determine it.

A sentence is considered to be relevant to an aspect, if at least two annotators considered it relevant. Sentences collected using keywords also can be irrelevant, for example a sentence mentioning Elon Musk ("Mask" in Russian spelling) is not relevant to the mask aspect. Multiple annotations for a relevant sentence are translated to three sentiment classes: positive, negative, and other (comprising neutral, contradictory, and unclear cases) using the following rules:

- a sentence has the positive score, if the number of positive annotations is more than the number of all other annotations for this sentence;
- a sentence has the negative score, if the number of negative annotations is more than the number of all other annotations for this sentence;
- otherwise the sentence is assigned to another category.

For example, the following sentence "the mask allows you not to maintain health, but to save your family budget" had three different labels from annotators: positive, negative, and impossible to determine. This sentence in fact need more context to precisely determine its sentiment towards masks, the attitude depends on interpretation. According to the above described rules, its resulting sentiment category is other.

Table 1 provides sizes of resulting categories for each aspect. It can be seen that the attitudes to masks and quarantine are mainly positive, the attitudes to government actions are mainly negative.

Aspect	Relevant	Negative	Positive	Other
Masks	5097	861 (17%)	1011 (20%)	3225 (63%)
Vaccines	2604	601 (23%)	538 (21%)	1465 (56%)
Quarantine	3515	244 (7%)	868 (25%)	2403 (68%)
Government	1585	1027 (65%)	54 (3%)	504 (32%)

Table 1. Sizes of sentiment categories

3.2 Analysing Annotators Disagreement

The most significant disagreement between annotators concerns assigning positive and negative scores to the same user's post. We found the following main cases for positive-negative disagreement between annotators.

First case. An author of a comment describes an opinion of another person, disagreement of the author with this opinion can be seen. In such cases some annotators can assess the sentiment of the author; other annotators can give label "positive and negative" (because two opinions are seen) or "impossible to determine"; the third annotator can select the described position because it takes most part of the sentence. For example (all examples are translated from Russian):

– *My father is so ... He endlessly repeats that masks, like vaccination, are a way of enslavement and he has an eternal "they are watching us" in his mind, I endlessly tease him, they say, be careful.*
– *But, just they think, since they are already sick, they no longer need a mask, there is nothing to defend against and they sneeze at everyone.*

Second case. The author tries to offend another participant of the dialogue using the aspect words:

– *well, nothing, nothing, someday for people like you they will definitely come up with vaccinations - from stupidity.* Here one annotator consider this comment as irrelevant to vaccines, other two annotators provide contradictory opinions (positive-negative)

Third case A comment describes some violations of mask or quarantine regimes. Some annotators consider such sentences as factual, neutral, other annotators try to infer some positive or negative positions. for example:

– *Because few tourists comply with the quarantine measures.*

Also typos may occur, which are difficult to explain. Because of all above-mentioned problems, we try to have at least three annotations for each comment.

4 Architecture and Methods for COVID Aspects Analysis

In the scope of this work, we use the RuBERT-conversational language model as a powerful feature extractor for classification. RuBERT-conversational is the BERT language model pre-trained on a large number of Russian tweets by the DeepPavlov project[2]. It greatly fits our needs because it was tuned on spoken and informal language data.

As the original BERT model, the input sequence of this model is either one or two sentences framed with special tokens:

$$[CLS], A_1, \ldots, A_m, [SEP], B_1, \ldots, B_n[SEP]$$

where $A_1, \ldots, A_m$ are tokens of the first sentence, $B_1, \ldots, B_n$ are tokens of the second sentence, $[SEP]$ is a special separating token, and $[CLS]$ is a special token, which represents the whole input sequence for classification tasks.

BERT returns hidden representations of every token of the input as the output. Furthermore, the representation of the $[CLS]$ token is processed by a fully connected layer, which was pre-trained for the Next Sentence Prediction objective, and the tanh activation function.

For the relevance determination and sentiment classification tasks, we added two fully-connected layers, containing 256 and K (a number of classes) outputs respectively, on top of the final representation of $[CLS]$. These layers are preceded by dropout layers with the rate of 0.5 and followed by the ReLU activation function:

$$\begin{aligned}
&H_1^d = dropout(0.5)(H_{[CLS]});\\
&H_2 = ReLU(W_1 H_1^d + b_1);\\
&H_2^d = dropout(0.5)(H_2);\\
&output = W_2 H_2^d + b_2;
\end{aligned}$$

where $H_{[CLS]} \in [-1, 1]^{768}$ is the embedding vector of $[CLS]$, $W_1 \in \mathbb{R}^{256\times768}$, $b_1 \in \mathbb{R}^{256}$, $W_2 \in \mathbb{R}^{K\times256}$, $b_1 \in \mathbb{R}^{K}$ are trainable parameters of the layers, and K is the number of outputs which is equal to number of classes in a task.

We formulate and solve the original classification tasks in different ways. First of all, we trained separate classifiers for each aspect. In that case, a document is an input to a classifier, and the output is either its relevance (0 or 1) or its sentiment (positive, negative, and other) to a considered aspect. In the second case, a document must be relevant to an aspect.

Secondly, the relevance determination problem was also postulated as a Natural Language Inference (NLI) problem. In that case, a classifier operates with all the given aspects and is able to learn relevance relations for new aspects if new data comes. The input of the classifier is a pair (s, h) of a sentence and an affirmative hypothesis about its relevance to an aspect, and the output is

[2] https://huggingface.co/DeepPavlov/rubert-base-cased-conversational

whether h is true (0 or 1). For example, h can state "Is relevant to masks" or "Is relevant to vaccines".

Thirdly, we formulate the sentiment classification problem as an NLI problem as well. In that case, for each triple (s, h) of a sentence and an affirmative hypothesis about its sentiment towards a relevant aspect, the classifier is trained to predict whether h holds truth (0 or 1). In that case, h may be "Is positive to masks" or "Is negative to quarantine".

Finally, the sentiment classification problem was stated as a Question Answering (QA) problem. In this formulation, we train a classifier to predict the sentiment polarity given a pair (s, a) of an expression and an aspect. In that case, a is simply an aspect, such as "Masks" or "Quarantine", and the output is a sentiment category. We decided not to use QA formulation for the relevance determination task because in that case it is equivalent to the NLI formulation.

Task	Epochs	LR
Sentiment Classification (NLI)	4	5e-6
Sentiment Classification (QA)	7	1e-5
Relevance Determination (NLI)	3	5e-5
RD and SC (aspect-specific)	7	1e-5

Table 2. Hyperparameters used by neural networks for different tasks.

5 Results of Experiments

During the experiments, we compare several variants of RuBERT-based models with classical machine learning methods, namely, Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB), Gradient Boosting (GB), and Random Forest (RF). Implementations of the classical algorithms were taken from the scikit-learn library[3]. These models receive tf-idf vectors as the inputs. To obtain the vectors, we tokenized and lemmatized texts, dropped stopwords, punctuation marks, and words that are seen less than in five documents. We tuned their hyperparameters with a Bayesian Optimization algorithm realized in the tune-sklearn library.

We utilized an implementation of the RuBERT-conversational model from the Transformers library. Other steps were performed with the PyTorch library [4]. The models were trained with the standard back-propagation algorithm. The size of a batch was set to 64. We utilized cross-entropy loss as a loss function, AdamW as an optimizer. OneCyclicLR with maximum learning rate of 3e-5 was utilised for learning aspect-specific SC and RD tasks (in the standard formulation). Other

[3] https://scikit-learn.ru/
[4] https://pytorch.org/

Aspect	Model	Accuracy	Precision	Recall	F1
Vaccines	SVM	99.45	98.47	99.23	98.85
	MNB	98.60	96.93	97.18	97.06
	BNB	98.45	95.17	98.46	96.79
	RF	**99.67**	**98.98**	**99.62**	**99.30**
	GB	99.64	**98.98**	99.49	99.23
Quarantine	SVM	98.27	**98.82**	95.73	97.25
	MNB	97.54	96.64	95.64	96.14
	BNB	97.81	96.76	96.39	96.58
	RF	98.30	97.61	**97.06**	97.34
	GB	**98.36**	98.36	96.49	**97.41**
Masks	SVM	99.09	98.64	99.41	99.02
	MNB	98.12	97.48	98.50	97.98
	BNB	98.63	98.43	98.63	98.53
	RF	99.12	98.64	99.48	99.06
	GB	**99.33**	**98.77**	**99.80**	**99.28**
Government	SVM	85.41	49.68	49.06	**49.37**
	MNB	75.78	32.94	**64.78**	43.67
	BNB	82.83	42.03	48.64	45.09
	RF	86.17	52.96	41.30	46.41
	GB	**87.21**	**65.05**	25.37	36.50
Average	SVM	95.55	86.40	85.86	**86.12**
	MNB	92.51	81.00	**89.03**	83.71
	BNB	94.43	83.10	85.53	84.25
	RF	95.82	87.05	84.37	85.53
	GB	**96.14**	**90.29**	80.29	83.10

Table 3. Performance of classical machine-learning methods in the relevance determination task

parameters were specific to the tasks and described in Table 2. In addition, we kept track of a current best (according to F1-score) model after each epoch.

All the models were tested with a random stratified train-test split, with a test size of 0.3. More precisely, the original texts were split into those collections, whereas the task-specific datasets were formed based on the same stratified train-test split.

Table 3 shows the performance of classical machine learning methods in the relevance determination task. The low results of classification for the "government actions" aspect can be explained with the diversity of lexical expressions of this aspect in comments. Some sentences do not contain direct mentions of this aspect, but nevertheless, express some opinion. Table 4 provides macro-averaged scores of classical machine-learning methods for the sentiment classification task.

Aspect	Model	Accuracy	macroPrec	macroRecall	macroF1
Vaccines	SVM	56.41	44.91	**44.29**	**44.27**
	MNB	54.49	42.52	44.02	43.10
	BNB	56.41	44.45	33.42	38.05
	RF	56.79	45.99	26.82	33.55
	GB	**58.59**	**54.58**	21.09	30.16
Quarantine	SVM	58.82	30.99	**45.46**	**35.94**
	MNB	60.82	29.54	39.94	33.92
	BNB	65.84	30.44	26.27	28.20
	RF	68.79	**43.54**	24.83	31.02
	GB	**69.73**	41.46	16.26	22.36
Masks	SVM	64.94	45.25	46.23	45.74
	MNB	61.81	41.60	**51.73**	**45.98**
	BNB	65.99	46.19	37.13	40.94
	RF	**67.69**	**56.64**	28.95	38.25
	GB	66.32	55.88	18.47	27.60
Government	SVM	**62.89**	**88.21**	35.68	**40.82**
	MNB	59.54	41.52	37.64	39.21
	BNB	**62.89**	36.51	36.16	36.33
	RF	61.22	42.08	40.73	40.47
	GB	61.84	33.60	**41.04**	36.95
Average	SVM	60.77	**52.34**	42.91	**41.69**
	MNB	59.16	38.80	**43.33**	40.55
	BNB	62.78	39.40	33.24	35.88
	RF	63.62	47.06	30.33	35.82
	GB	**64.12**	46.38	24.21	29.27

Table 4. Performance of classical machine learning methods in the sentiment classification task

Table 5 and Table 6 compare results of the best (according to F1-score) classical methods to RuBERT-based models in relevance determination and sentiment classification tasks correspondingly.

Aspect	Model	Accuracy	Precision	Recall	F1
Vaccines	NLI	**99.70**	**98.98**	**99.74**	**99.36**
	RuBERT	99.67	**98.98**	99.62	99.30
	RF	99.67	**98.98**	99.62	99.30
Quarantine	NLI	98.27	98.16	96.39	97.27
	RuBERT	**98.51**	97.99	**97.34**	**97.67**
	GB	98.36	**98.36**	96.49	97.41
Masks	NLI	99.27	98.77	99.67	99.22
	RuBERT	**99.48**	**99.09**	**99.80**	**99.45**
	GB	99.33	98.77	**99.80**	99.28
Government	NLI	**88.45**	**65.59**	42.77	51.78
	RuBERT	86.84	54.55	**55.35**	**54.94**
	SVM	85.41	49.68	49.06	49.37
Average	NLI	**96.42**	**90.38**	84.64	86.91
	RuBERT	96.13	87.65	**88.03**	**87.84**
	SVM	95.55	86.40	85.86	86.12

Table 5. The results of RuBERT-based models and their comparison with the best classical methods in the relevance determination task

As we can see from the tables, both classical methods and neural networks determine the relevance of messages with the high quality when the messages include direct mentions of an aspect. In more complex scenarios, neural networks show better results. In the sentiment classification task, neural networks also achieve higher scores, because they consider context and the order of words.

Finally, the use of the NLI and the QA formulations increased the scores of the sentiment classification task, whereas the QA formulation performs slightly better. It may be explained by the introduction of additional aspect-related features to the input of the models and by the usage of the whole collection of sentences for training. The lowest results of macro measures are obtained for the government aspect, which can be explained with small number of examples in the positive class.

As for the RD task, the introduction of the new formulations did not increase overall quality. This behavior may be caused by the fact that the task is too 'simple' for the model to improve further.

Aspect	Model	Accuracy	MacroPrec	MacroRecall	MacroF1
Vaccines	NLI	**66.67**	**63.88**	59.74	61.24
	QA	66.54	63.62	**61.62**	**62.40**
	RuBERT	62.56	58.86	59.76	59.25
	SVM	56.41	44.91	44.29	44.27
Quarantine	NLI	**74.38**	**70.07**	50.82	53.52
	QA	73.81	63.64	**59.38**	**61.13**
	RuBERT	73.72	57.74	54.64	55.85
	SVM	58.82	30.99	45.46	35.94
Masks	NLI	70.83	**64.29**	56.76	59.21
	QA	**71.03**	63.83	**61.76**	**62.69**
	RuBERT	65.27	57.84	56.14	56.14
	MNB	61.81	41.60	51.73	45.98
Government	NLI	**69.81**	44.79	41.51	41.04
	QA	68.97	43.19	42.17	41.91
	RuBERT	68.76	43.84	**46.21**	**44.83**
	SVM	62.89	**88.21**	35.68	40.82
Average	NLI	**70.42**	**60.76**	52.21	53.75
	QA	70.09	58.57	**56.23**	**57.03**
	RuBERT	67.58	54.57	54.19	54.02
	SVM	60.77	52.34	42.91	41.69

Table 6. The results of RuBERT-based models and their comparison with the best classical methods in the sentiment classification task

6 Conclusion

In this paper, we introduce a specialized Russian dataset of Russian users' comments about COVID-19 aspects. The dataset contains sentences with sentiment scores towards four topics widely discussed such as masks, vaccines, quarantine, and government measures. Each comment is scored by three annotators on average.

We studied approaches to aspect-based sentiment analysis of the created dataset. We solved two tasks, namely Relevance Determination (RD), which aims to predict whether a sentence is relevant to an aspect of the pandemic, and Sentiment Classification (SC), which classifies the sentiment expressed towards an aspect in a sentence.

We applied and tested various methods of machine learning, including fine-tuning of the pre-trained RuBERT model. The best results were obtained by RuBERT model in special settings called Natural Language Inference (NLI) and Question Answering (QA), in which an additional sentence is added to a classified sentence, indicating a target aspect.

The created collection is publicly available[5].

Acknowledgements. The reported study was funded by RFBR according to the research project № 20-04-60296.

References

1. Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., Shah, Z.: Top concerns of tweeters during the covid-19 pandemic: infoveillance study. Journal of medical Internet research **22**(4), e19016 (2020)
2. Alamoodi, A., Zaidan, B., Zaidan, A., Albahri, O., Mohammed, K., Malik, R., Almahdi, E., Chyad, M., Tareq, Z., Albahri, A., et al.: Sentiment analysis and its applications in fighting covid-19 and infectious diseases: A systematic review. Expert systems with applications p. 114155 (2020)
3. Barkur, G., Vibha, Kamath, G.: Sentiment analysis of nationwide lockdown due to covid 19 outbreak: Evidence from india. Asian Journal of Psychiatry **51** (Jun 2020). https://doi.org/10.1016/j.ajp.2020.102089
4. Chandrasekaran, R., Mehta, V., Valkunde, T., Moustakas, E.: Topics, trends, and sentiments of tweets about the covid-19 pandemic: Temporal infoveillance study. Journal of medical Internet research **22**(10), e22624 (2020)
5. De Santis, E., Martino, A., Rizzi, A.: An infoveillance system for detecting and tracking relevant topics from italian tweets during the covid-19 event. IEEE Access **8**, 132527–132538 (2020)
6. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR **abs/1810.04805** (2018), http://arxiv.org/abs/1810.04805
7. Huang, X., Smith, M.C., Paul, M.J., Ryzhkov, D., Quinn, S.C., Broniatowski, D.A., Dredze, M.: Examining patterns of influenza vaccination in social media. In: Workshops at the thirty-first AAAI conference on artificial intelligence (2017)

[5] https://github.com/LAIR-RCC/RussianCovidDataset

8. Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., Ali, A., Sheikh, A.: Artificial intelligence–enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study. Journal of medical Internet research **23**(4), e26627 (2021)
9. Hutto, C., 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. vol. 8 (2014)
10. Karimi, A., Rossi, L., Prati, A., Full, K.: Adversarial training for aspect-based sentiment analysis with BERT. CoRR **abs/2001.11316** (2020), https://arxiv.org/abs/2001.11316
11. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.: NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 437–442. Association for Computational Linguistics, Dublin, Ireland (Aug 2014). https://doi.org/10.3115/v1/S14-2076, https://www.aclweb.org/anthology/S14-2076
12. Kuratov, Y., Arkhipov, M.: Adaptation of deep bidirectional multilingual transformers for russian language. CoRR **abs/1905.07213** (2019), http://arxiv.org/abs/1905.07213
13. Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., et al.: Textblob: simplified text processing. Secondary TextBlob: simplified text processing **3** (2014)
14. Miao, Z., Li, Y., Wang, X., Tan, W.: Snippext: Semi-supervised opinion mining with augmented data. CoRR **abs/2002.03049** (2020), https://arxiv.org/abs/2002.03049
15. Ovchinnikova, I., Ermakova, L., Nurbakova, D.: Sentiments in russian medical professional discourse during the covid-19 pandemic. In: Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media. pp. 99–108 (2020)
16. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. CoRR **abs/1802.05365** (2018), http://arxiv.org/abs/1802.05365
17. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al.: Semeval-2016 task 5: Aspect based sentiment analysis. In: International workshop on semantic evaluation. pp. 19–30 (2016)
18. Rietzler, A., Stabinger, S., Opitz, P., Engl, S.: Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification. CoRR **abs/1908.11860** (2019), http://arxiv.org/abs/1908.11860
19. 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. Association for Computational Linguistics, Santa Fe, New Mexico, USA (Aug 2018), https://www.aclweb.org/anthology/C18-1064
20. Sanders, A.C., White, R.C., Severson, L.S., Ma, R., McQueen, R., Alcântara Paulo, H.C., Zhang, Y., Erickson, J.S., Bennett, K.P.: Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of covid-19 twitter discourse. medRxiv (2021). https://doi.org/10.1101/2020.08.28.20183863, https://www.medrxiv.org/content/early/2021/03/20/2020.08.28.20183863

21. Sharma, K., Seo, S., Meng, C., Rambhatla, S., Dua, A., Liu, Y.: Coronavirus on social media: Analyzing misinformation in twitter conversations. CoRR **abs/2003.12309** (2020), https://arxiv.org/abs/2003.12309
22. Song, Y., Wang, J., Jiang, T., Liu, Z., Rao, Y.: Attentional encoder network for targeted sentiment classification. CoRR **abs/1902.09314** (2019), http://arxiv.org/abs/1902.09314
23. Sun, C., Huang, L., Qiu, X.: Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 380–385 (2019)
24. Tang, D., Qin, B., Feng, X., Liu, T.: Effective LSTMs for target-dependent sentiment classification. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 3298–3307. The COLING 2016 Organizing Committee, Osaka, Japan (Dec 2016), https://www.aclweb.org/anthology/C16-1311
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. CoRR **abs/1706.03762** (2017), http://arxiv.org/abs/1706.03762
26. Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D., Bengio, Y.: Manifold mixup: Better representations by interpolating hidden states. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 6438–6447. PMLR (09–15 Jun 2019), http://proceedings.mlr.press/v97/verma19a.html
27. Wang, Y., Huang, M., Zhu, X., Zhao, L.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 606–615. Association for Computational Linguistics, Austin, Texas (Nov 2016). https://doi.org/10.18653/v1/D16-1058, https://www.aclweb.org/anthology/D16-1058
28. Yang, H., Zeng, B., Yang, J., Song, Y., Xu, R.: A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction. CoRR **abs/1912.07976** (2019), http://arxiv.org/abs/1912.07976
29. Zhao, P., Hou, L., Wu, O.: Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. CoRR **abs/1906.04501** (2019), http://arxiv.org/abs/1906.04501
30. Zhuravlev, A., Kitova, D.: Emotional characteristics of the attitude of social network users towards the coronavirus infection. Proceedings of the 'V.M. Behterev and modern personality psychology' conference pp. 208–2011 (2020)