Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 190 (2021) 637–642

Procedia Computer Science

www.elsevier.com/locate/procedia

2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society

Data-Driven Model for Emotion Detection in Russian Texts

Alexander Sboev[a,b,1], Aleksandr Naumov[a], Roman Rybka[a]

[a]*National Research Centre "Kurchatov Institute", Moscow, Russia*
[b]*National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia*

Abstract

An important task in the field of automatic data analysis is detecting emotions in texts. The paper presents the approach of emotion recognition for text data in Russian. To conduct an emotion analysis, a method was created based on vector representations of words obtained by the ELMo language model, and subsequent processing by an ensemble classifier. To configure and test the created method, a specially prepared dataset of texts for five basic emotions – joy, sadness, anger, fear, and surprise – is used. The dataset was prepared using a crowdsourcing platform and a home-grown procedure for collecting and controlling annotators' markup. The overall accuracy is 0.78 (by the F1-macro score), which is currently the new state of the art for Russian. The results can be used for a wide range of tasks, for example: monitoring social moods, generating control signals for mobile robotic systems, etc.

Keywords: text analysis ; natural language processing ; emotion detection ; machine learning ; crowdsourcing

1. Introduction

The problems of extracting automatically the emotions of the author manifested in their text are in high demand for a wide range of practical applications. The latter include, in particular, determining sources of threats of various types or the presence of illegal context, consulting, and marketing systems, as well as a prospective application in the control interfaces of robotic devices as a means of determining the priority of a command. The problem under consideration has been extensively studied in the world; this is evidenced by the presence of labeled data samples, as well as some works on methods.

* Corresponding author. Tel.: +7-926-253-7217
E-mail address: Sboev_AG@nrcki.ru

10.1016/j.procs.2021.06.075

In recent years, the development of emotion analysis methods has been facilitated by competitions organized within thematic conferences. Of especial interest is the WASSA-2017 [9] competition, for which one of the most representative datasets was collected. That dataset contains 7097 messages from the Twitter social network with labeling of four emotions – anger, fear, joy, and sadness – and their intensity magnitudes.

The best accuracy in the competition, 0.75 by Pearson's correlation [9], was shown by a solution based on an ensemble of neural network models that solves the multi-label classification problem of predicting intensities of different emotions [3]. Each model was configured to determine the intensity of individual emotion. Vector representations of words and dictionaries of emotive vocabulary were used as input features for the model.

Later, during the SemEval-2018 [8] competition (Task 1), the WASSA-2017 competition dataset was expanded to 22,000 examples for three languages: Arabic, English, and Spanish. The best solution for the English language consisted of a complex of binary classifiers based on Gradient Boosting and Random Forest [1]. The authors applied modern deep learning models pre-trained on large corpora of third-party data to extracting features of texts. For instance, as the features they used vectors obtained from the DeepMoji [2] and Sentiment neurons [12] models, as well as a vector representation of words from Skip thoughts [5]). This solution, combined with the use of dictionaries of emotional vocabulary made it possible to achieve a better result with a Pearson correlation coefficient of 0.8 to determine the intensity of various emotions.

Regarding Russian, we should mention the statistical part-time approach [14], where the emotion analysis problem was considered as a task of scoring psycholinguistic indicators. The most developed methods are based on rich dictionaries of emotive vocabulary [7, 10] and their application in retrospect.

Unfortunately, the modern machine learning methods for the English language cannot be applied to similar tasks of extracting emotions in a text in Russian because there are no similar sets of labeled data. Therefore, the task of creating and developing tools for recognizing the emotions of the author from texts written in Russian is relevant and is the goal of this work.

In this work, to implement the data-driven model for emotion detection, a special corpus of 9 668 sentences was formed, labeled with five basic emotions: joy, sadness, anger, fear, and surprise. Section 2 describes the methods used for crowdsource-collecting, labeling, and verifying the data. The model for emotion detection in texts, described in Section 3, is based on an ensemble of several classifiers which receive ELMo vector representation of words [11] as input features. The performance of the proposed ensemble classifier is presented in Section 5. Compared to the existing basic methods (listed in Section 4) used to solve such problems, the obtained results justify the effectiveness of the created solution.

2. Emotive sentence dataset

For the formation of the corpus, we collected sentences in Russian from several sources: posts of the LiveJournal social network[1], texts of the online news agency Lenta.ru[2] and Twitter microblog posts [13].

Only those sentences were selected that contained marker words from the dictionary of the emotive vocabulary of the Russian language [15]. We manually formed a list of marker words for each emotion by choosing words from different categories of the dictionary. As the result, for example, the list for the *anger* emotion contained such words as "anger", "furious", "to annoy", etc.; the list for the *fear* emotion contained "fear", "horror", "creepy", etc.

In total, 3107 sentences were selected from LiveJournal posts, 3131 sentences from Lenta.Ru, and 3430 sentences from Twitter. After selection, sentences were offered to annotators for labeling.

2.1. Annotation procedure

Annotating sentences with labels of their emotions was performed with the help of a crowdsourcing platform[3]. Only those of the 30% of the best-performing active users (by the platform's internal rating) who spoke Russian and were over 18 years old were allowed into the annotation process.

[1] Rusprofiling corpus: http://rusprofilinglab.ru/rusprofiling-atpan/corpus/
[2] Lenta.Ru corpus: https://github.com/yutkin/Lenta.Ru-News-Dataset
[3] Yandex.Toloka: https://yandex.ru/support/toloka/index.html?lang=en

Before a platform user could be employed as an annotator, they underwent a training task, after which they were to mark 25 trial samples with more than 80% agreement compared to the annotation that we had performed ourselves. Upon successful completion of the training task and this trial run, the user was allowed to take the main tasks, each task consisting of 10 sentences for annotation, one of which was a control one that we had labeled. For this additional control, we labeled 200 sentences. If the accuracy of an annotator on the control sentences (including the trial run) became less than 70%, or if the accuracy was less than 66% over the last six control samples, the annotator was dismissed. As additional checks, completing a 10-sentence task in less than 30 seconds or assigning more than eight identical labels in a row was considered suspicious, causing the labeled sentences to be checked manually and excluded from the dataset if proven unreliable.

Sentences were split into tasks and assigned to annotators so that each sentence was annotated at least three times.

2.2. Rules of assigning emotion labels

The annotators' task was: "What emotions did the author express in the sentence?". The annotators were allowed to put an arbitrary number of the following emotion labels:

Joy: including pleasure, kindness, friendship. Emoticons such as :), :D, xD, :P or closing brackets ")))" can also be markers of joy;

Anger: including aggression, cruelty, hatred, insults;

Sadness: including sadness, grief, pity, loneliness, tears, as well as related emoticons, such as ":(" or opening brackets "((((";

Fear: including anxiety (except when the Russian equivalents of "terrible", "terrific" and alike are used as emphasizing modifiers in a different context, for example, "I'm terribly sorry" being *Sadness* emotion group);

Surprise: including messages where the author expresses impressions of some unusual, unexpected or strange object, event, or phenomenon. Markers of surprise can be emoticons, for example, emoticons "O_o, o_o, Oo, 0_0", as well as the words "Oooooh", "wow", "go crazy", etc.

Sentences which the annotators believed to express no author's emotion were to be assigned *No emotion* label.

The final label for the sentence was *No emotion* if less than 60% annotators put any emotion labels on it (regardless of their types). A label of a specific emotion was assigned to a sentence if put by more than half of the annotators (except those who put *No emotion*).

After such aggregation of annotators' labels, the corpus contains 9668 sentences, 5609 of which contain one or more emotions, and 4059 sentences contain none.

The number of labels by data source is presented in Table 1.

Table 1. The number of emotion labels in different subsets of our dataset.

Data source	Joy	Sadness	Fear	Anger	Surprise	No emotions	Total sentences
Twitter	1233	1372	362	198	221	129	3430
Lenta.ru	190	92	126	121	195	2411	3131
LiveJournal	433	298	264	231	398	1519	3107
Total	1856	1762	752	550	814	4059	9668

As can be seen from the table, the largest number of emotive samples is contained in the texts of micro-blogs messages, while sentences from news articles rarely reflect any emotion, even if they contain emotive words.

3. The ensemble classifier for emotion detection in sentences

The method was developed for solving the multi-label classification problem when each analysed sentence can be assigned several different emotion labels [16, 4]

For encoding a sentence we use vector representations of words obtained with the pretrained ELMo model[4]. Embedding Language Model (ELMo) [11] is a deep neural network model which involves two parts, a forward one which predicts the probability of t-th word w_t given the previous $1 \ldots t-1$ words of the sentence (equation 1), and a backward one which predicts the probability of w_t given the subsequent $t+1 \ldots N$ words (equation 2):

$$p(w_1, w_2, \ldots, w_N) = \prod_{t=1 \ldots N} p(w_t | w_1, w_2 \ldots, w_{t-1}) \tag{1}$$

$$p(w_1, w_2, \ldots, w_N) = \prod_{t=1 \ldots N} p(w_t | w_{t+1}, w_{t+2}, \ldots, w_N) \tag{2}$$

The vector representation of a word is formed by extracting the internal state of the model as the combination of representations from the outputs of the forward and backward intermediate layers of the bidirectional language model. The vector representation for the whole sentence is obtained by averaging the resulting word vectors.

The sentences thus encoded are processed with classifiers (see Figure 1), among which we test a set of supervised machine learning methods[5], including the support vector machine (SVM) with linear kernel, gradient boosting (XGB), stochastic gradient descent, and logistic regression. Search for the hyperparameters of the models was performed with an AutoML method based on the TPOT library [6]. After training and testing the classifiers, the best ones (according to the F1-macro score) are selected for inclusion into the ensemble.

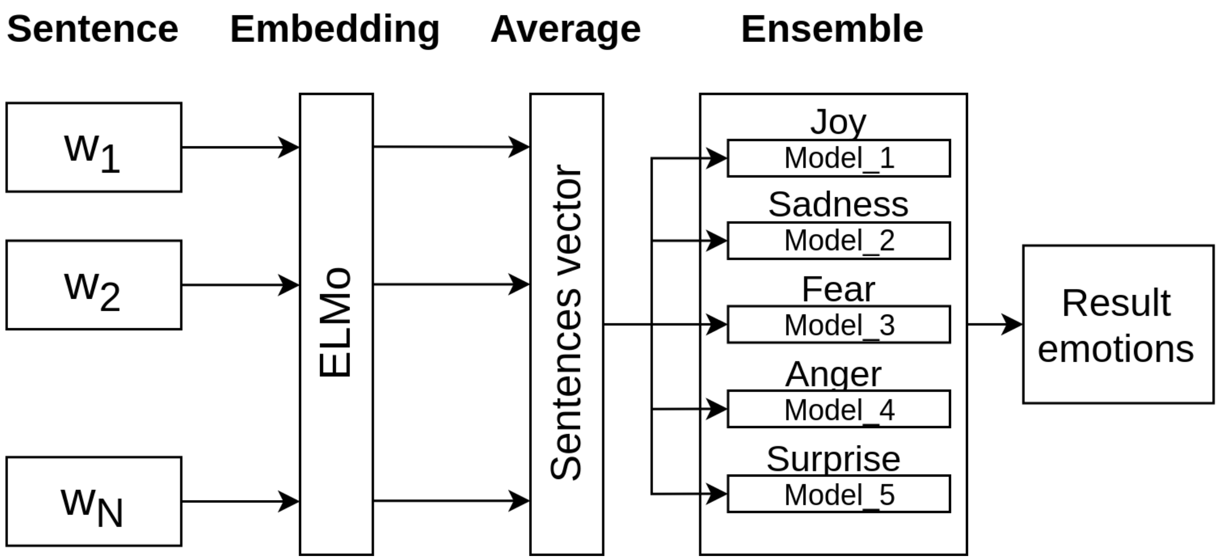

Fig. 1. The scheme of the model for emotion recognition in a sentence

4. Experiments

4.1. Baselines for comparison

While evaluating the proposed solution, we compare it against several basic estimates:

1. Random: The emotion label is chosen randomly for each sentence;

[4] ELMo model: docs.deeppavlov.ai/en/master/features/pretrained_vectors.html#elmo
[5] Python library scikit-learn. url: https://scikit-learn.org

2. SVM (TF-IDF): The classifier based on SVM with linear kernel, where the input sentence is encoded with the TF-IDF frequency method using 4-to-8-long character n-grams;
3. Lexicon: The classifier based on dictionaries of emotive vocabulary. The emotion label is determined by the presence of words from the emotive vocabulary for the corresponding emotion. The vocabulary is the same that was used when pre-selecting sentences for annotation as described in section 3.2. Dictionary lookup is performed for lower-case lemmas of words.

4.2. Accuracy metrics

The accuracy of the obtained models is measured with the F1 metric:

$$\begin{aligned} P &= \frac{TP}{TP+FP}, \\ R &= \frac{TP}{TP+FN}, \\ F1 &= \frac{2*P*R}{P+R}). \end{aligned} \tag{3}$$

Here TP is the number of true positive estimations, FP is the number of false-positive estimations, FN is the number of false-negative estimations.

F1 is calculated in two variations, F1-macro and F1-micro. In the first case, the measure is calculated for each class separately and then averaged. in F1-micro, the calculation is carried out for all examples together. This allows one to evaluate the results in the case of imbalanced classes in the data.

5. Results

After the adjustment and comparison of the classifiers, the final solution is an ensemble of the following five binary classifiers:

1. for the *joy* emotion: a model based on a support vector machine with a linear kernel;
2. for the *sadness* emotion: a model based on logistic regression;
3. for the *fear* emotion: a model based on stochastic gradient descent with PCA preprocessing of input features;
4. for the *anger* emotion: a model based on logistic regression;
5. for the *surprise* emotion: a model based on gradient descent and logistic regression methods, with preliminary normalization of the data: the ratio of the difference in value and its mean to standard deviation.

The F1-scores of the selected classifiers in comparison with the results of the baseline methods are presented in Table 2.

Table 2. The F1-micro (mic.) and F1-macro (mac.) of detecting different emotions.

Model	Joy		Sadness		Fear		Anger		Surprise		Mean	
	mic.	mac.	mic.	mac.	mic.	mac.	mic.	mac.	mic.	mac.	mic.	mac.
Random	0.5	0.45	0.52	0.45	0.51	0.39	0.5	0.38	0.5	0.4	0.51	0.41
SVM (TF-IDF)	0.87	0.71	0.88	0.72	0.94	0.68	**0.95**	0.53	**0.93**	0.69	0.91	0.66
Lexicon	0.76	0.65	0.7	0.59	0.79	0.65	0.76	0.62	0.79	0.64	0.76	0.63
Our ensemble	**0.93**	**0.88**	**0.91**	**0.83**	**0.95**	**0.75**	0.92	**0.66**	**0.93**	**0.77**	**0.93**	**0.78**

When training and testing on 80% and 20% of the dataset, the developed method demonstrates the F1-macro of 0.78, which exceeds random choice, the basic SVM classifier, and the dictionary method by 0.37, 0.12, and 0.15, respectively.

Conclusion

The paper presents a study of methods for emotion analysis of text data in Russian. The formed corpus of examples consists of sentences from several sources (blogs, microblogs, news), which allows creating methods to analyse various types of texts. The created methodology for building the dataset based on applying a crowdsourcing service can be used to expand the number of examples to improve the accuracy of supervised classifiers.

Based on the assembled corpus, the method for solving the emotion analysis problem has been proposed and validated. At its core are vector representations of sentences obtained by the ELMo deep learning language model and the ensemble of binary classifiers trained using an AutoML algorithm. The obtained accuracy is the new baseline for this type of task in Russian.

The future work will be aimed at 1) expanding the corpus, 2) improving the methodology of forming a dataset with the help of crowdsourcing, 3) developing more efficient methods of multi-language emotion analysis.

The dataset can be achieved by requesting through https://sagteam.ru/en/ website.

Acknowledgements

The reported study was funded by an internal grant of the NRC ”Kurchatov Institute” (Order No. 1359) and has been carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC “Kurchatov Institute”, http://ckp.nrcki.ru/.

Thanks to Alexei Serenko for informative comments and suggestions in the process of preparing the text of the article.

References

[1] Duppada, V., Jain, R., Hiray, S., 2018. Seernet at semeval-2018 task 1: Domain adaptation for affect in tweets. arXiv preprint arXiv:1804.06137 .

[2] Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., Lehmann, S., 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. arXiv preprint arXiv:1708.00524 .

[3] Goel, P., Kulshreshtha, D., Jain, P., Shukla, K.K., 2017. Prayas at emoint 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 58–65.

[4] Jabreel, M., Moreno, A., 2019. A deep learning-based approach for multi-label emotion classification in tweets. Applied Sciences 9, 1123.

[5] Kiros, R., Zhu, Y., Salakhutdinov, R.R., Zemel, R., Urtasun, R., Torralba, A., Fidler, S., 2015. Skip-thought vectors, in: Advances in neural information processing systems, pp. 3294–3302.

[6] Le, T.T., Fu, W., Moore, J.H., 2020. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. Bioinformatics 36, 250–256.

[7] Loukachevitch, N., Levchik, A., 2016. Creating a general russian sentiment lexicon, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16), pp. 1171–1176.

[8] Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S., 2018. Semeval-2018 task 1: Affect in tweets, in: Proceedings of the 12th international workshop on semantic evaluation, pp. 1–17.

[9] Mohammad, S.M., Bravo-Marquez, F., 2017. Wassa-2017 shared task on emotion intensity. arXiv preprint arXiv:1708.03700 .

[10] Panchenko, A., 2018. Sentiment index of the russian speaking facebook. arXiv preprint arXiv:1808.07851 .

[11] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365 .

[12] Radford, A., Jozefowicz, R., Sutskever, I., 2017. Learning to generate reviews and discovering sentiment. arXiv preprint arXiv:1704.01444 .

[13] Rubtsova, Y., 2012. Avtomaticheskoye postroyeniye i analiz korpusa korotkikh tekstov (postov mikroblogov) dlya zadachi razrabotki i trenirovki tonovogo klassifikatora. inzheneriya znaniy i tekhnologii semanticheskogo veba. Inzheneriya znanij i tekhnologii semanticheskogo veba 1, 109–116.

[14] Sboev, A., Gudovskikh, D., Rybka, R., Moloshnikov, I., 2015. A quantitative method of text emotiveness evaluation on base of the psycholinguistic markers founded on morphological features, in: Sloot, P., Klimentov, A., Athanassoulis, G., Boukhanovsky, A. (Eds.), 4th International Young Scientist Conference on Computational Science (YSC), Elsevier B.V.. pp. 307–316. doi:10.1016/j.procs.2015.11.036.

[15] Shvedova, N., . Information retrieval system ”emotions and feelings in lexicographical parameters: Dictionary emotive vocabulary of the russian language”. URL: http://lexrus.ru/default.aspx?p=2876.

[16] Zhang, M.L., Zhou, Z.H., 2013. A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering 26, 1819–1837.