Emotion Classification in Russian: Feature Engineering and Analysis

Marina Kazyulina, Aleksandr Babii(✉), and Alexey Malafeev

National Research University Higher School of Economics, Nizhny Novgorod, Russia
mskazyulina@edu.hse.ru

Abstract. In this paper, we address the issue of identifying emotions in Russian informal text messages. For this purpose, a new large dataset of text messages from the most popular Russian messaging/social networking services (Telegram, VK) was compiled semi-automatically. Emojis contained in the text messages were used to annotate the data for emotions expressed. This paper proposes an integrated approach to text-based emotion classification combining linguistic methods and machine learning. This approach relies on morphological, lexical, and stylistic features of the text. Furthermore, the level of expressiveness was considered as well. As a result, an emotion classification model demonstrating near-human performance was designed. In this paper, we also report on the importance of different linguistic features of the text messages for the task of automatic emotive analysis. Additionally, we perform error analysis and discover ways to improve the model in the future.

Keywords: Machine learning · Emotion identification · Emotiveness · Sentiment analysis · Natural language processing

1 Introduction

This paper focuses on the development of an integrated approach to detecting emotions expressed in short informal texts, combining linguistic analysis and machine learning. We propose an automatic classifier of text messages based on which emotions are explicitly or implicitly present in the text. Some of real-world applications may include partial automation of linguistic text analysis and psychological testing in such fields as medicine, forensics, and HR management. Moreover, it may provide new capabilities in marketing and especially in targeting. Finally, it may also greatly assist further advancement of suicide prevention mechanisms on social media platforms such as Facebook.

This paper is a much extended and enhanced version of [10]. In comparison with [10], the present paper features additional feature construction and selection with tree-based ensemble models. It also includes feature importance analysis, in particular, analysis of: lexical markers, different types of emotives[1], as well as the use of parts of speech for determining emotion. We also perform error analysis and investigate the reasons

[1] By emotives we mean any language units, not only lexis, that are used to express emotions.

W. M. P. van der Aalst et al. (Eds.): AIST 2020, LNCS 12602, pp. 135–148, 2021.
https://doi.org/10.1007/978-3-030-72610-2_10

for most frequent misclassifications. Another contribution of this paper is a substantial review of existing emotion classifiers with feature comparison, as well as links to two manually anonymized datasets[2] and our custom data parser.

The paper is structured as follows: Sect. 2 overviews some recent research in the field of automatic emotion analysis and compares current work to some of the better-known papers; Sect. 3 describes the process of dataset creation; Sect. 4 gives details on the development of our model; Sect. 5 discusses results achieved; finally, Sect. 6 draws conclusions and outlines future work.

2 Related Work

Researchers have achieved good results on image-based emotion recognition [11]. However, classifying textual dialogues based on emotions is a relatively new research area. Another challenge for automated emotion detection is that emotions are complex concepts with fuzzy boundaries and with individual variations in expression and perception.

Emotiveness as an immanent function of language is carried out by the aid of emotive features on different language levels. A number of studies have found that lexical emotive features (dictionaries containing expressive words) are effective in automatic text analysis [14]. Additionally, morphological emotives have been successfully used in the development of a system for predicting the emotiveness of texts [6], which proves that psycholinguistic markers indicate strong emotional reactions of the text author. Among such psycholinguistic markers are the Trager coefficient (ratio of verb count to adjective count, with the normal value close to 1), the coefficient of certainty of action (verb to noun ratio, with the normal value also close to 1), and also the coefficient of aggressiveness (verbs to words in the text overall, with the normal value lower than 0.6). Other indicators at the punctuation and graphic levels, such as exclamation marks, question marks and uppercase words, can also help indicate expressed emotions in the text.

Machine learning algorithms have proven to be highly effective in terms of detection, classification, and quantification of emotions of text in any form [5]. Table 1 shows a number of most relevant studies that take advantage of various machine learning technique to solve the problem of text-based emotion classification for English. While a lot of progress has been made in this field, research into emotion detection in Russian-language texts is still extremely limited. The present work is aimed at addressing this problem for Russian, and some features that we use are language-specific (see Sect. 4).

To develop an automatic classifier of emotions expressed in text messages, it is necessary to adopt a classification of emotions. As of now, there is no generally accepted emotion classification. Ekman's classification [4] appears to be the most widely used. It includes happiness, surprise, sadness, anger, fear, disgust, and contempt. At the current stage of work and given the limitations of our data described in Sect. 3, we removed the categories of disgust and contempt for lack of appropriate data. Moreover, the categories of fear and surprise were merged into one category of uncertainty due to certain similarity of the ways in which these emotions are usually expressed.

[2] One with manually tagged messages, another with 50K messages marked up semi-automatically; for more details, please refer to Sect. 3.

Thus, the classifier presented in this paper recognizes five categories of text, according to the emotion most vividly expressed in it: happiness, sadness, aggression, uncertainty, and neutrality (no emotion).

3 Data

Our work is focused on the informal Internet-based discourse of personal messaging. For the purposes of our research, we compiled a new corpus of text messages. The compilation process included 4 stages: 1) Sourcing and parsing of data; 2) Data preprocessing; 3) Data processing; 4) Corpus annotation.

3.1 Parsing and Processing

Working with user data involves addressing the issue of the ethical use of personal information. According to the General Data Protection Regulation (GDPR), all companies operating in the European Union are obliged to provide their users with personal data in the possession of these companies upon request. Personal data includes not only the messages written by user but also the ones they received. This allowed us to request access to and use personal messages and public chats from VKontakte[3] (a social network often used as a messenger) and Telegram[4] (a messenger with some social network features) to create our message corpus. We chose these services because they are highly popular among Russian-speaking people.

A custom parsing algorithm[5] was developed to accommodate the Telegram and VK reporting formats. Regarding the parsed data, statistical population characteristics include people that are 12–35 y.o. with education level from general to postgraduate. The statistical population is represented by the sample of 4584 people. Statistically, most of the parsed messages are 1–9 words long and the average length is 4–5 words. In order to prepare the messages for further work, it is necessary to ensure that they are all anonymized and of the same length. In order to do so, all the messages consisting of more than 9 words were discarded. We also removed tags, hashtags, and messages with no Cyrillic symbols. Furthermore, e-mail addresses and phone numbers were replaced with special tokens <email> and <phone>, respectively. Then the corpus was normalized with regular expressions and lemmatized with the *rnnmorph*[6] lemmatizer. The final dataset of lemmatized messages consists of 1,800,000 messages.

3.2 Corpus Annotation

In order to annotate the messages with emotion labels automatically, it was decided to use emoji ideograms as an objective emotion indicator. Obviously, not all the messages in corpus contain emojis. Thus, it was decided to build 5 classifiers (one for each emotion)

[3] https://vk.com.
[4] https://telegram.org.
[5] https://github.com/asbabiy/AIST/blob/master/Parser.py.
[6] https://github.com/IlyaGusev/rnnmorph.

Table 1. Comparison of existing classifiers

Authors	Emotion classification used	Num. of classes	Rationale behind training data	Data	Precompiled emotion lexicons	Method	F1
Klinger 2018 (Baseline) [9]	Sadness, Fear, Joy, Love, Surprise, Anger	6	Hashtags	CrowdFlower, blogs, emotion corpora	–	Logistic regression	0.258
Seyeditabari et al. 2019 [15]	Sadness, Fear, Joy, Love, Surprise, Anger	6	Pretrained word embeddings	CrowdFlower	–	biGRU	0.632
Hasan et al. 2018 [8]	Happy-Active; Happy-Inactive; Unhappy-Active; Unhappy-Inactive	4	Keywords, multiple binary classifiers for labeling	Twitter	+	SVM	0.9
Canales et al. 2016 [2]	Anger, Disgust, Fear, Joy. Sadness, Surprise, Neutral	7	Pretrained word embeddings; WordNet and Oxford synsets	Alm and Aman corpora	+	SVM	0.745
Mohammad, Kiritchenko 2015 [12]	Anger, Disgust, Fear, Joy. Sadness, Surprise	6	Hashtags	Facebook posts	+	SVM	0.595
Chatterjee et al. 2019 [3]	Happy, Sad, Angry, Others	4	Top unigrams; small sets of emojis for each emotion	SemEval-2019	–	?	0.796

(*continued*)

Table 1. (*continued*)

Authors	Emotion classification used	Num. of classes	Rationale behind training data	Data	Precompiled emotion lexicons	Method	F1
Gupta et al. 2017 [7]	Happy, Sad, Angry, Others	4	Top n-grams; emoticons; sentence embeddings	Twitter	–	SS-LSTM	0.713
Zahiri, Choi 2017 [16]	Sad, Mad, Scared, Powerful, Peaceful, Joyful, Neutral	7	Word embeddings	TV show Friends transcript	–	SCNN	0.393
Abdul-Mageed, Ungar 2017 [1]	Anger, Disgust, Fear, Joy. Sadness, Surprise	6	Hashtags; synsets	Twitter	+	GRNNs	0.801
This study	Happiness, Sadness, Anger, Uncertainty, Neutral	5	Manually selected sets of emoji; bootstrap for labeling	VK, Telegram	–	Logistic regression	0.718

trained on those objectively labeled messages. Those classifiers then labeled the rest of the data. The details of the process are given below.

First, the emojis that were used less than 40 times overall were removed from the observed data. The rest of the emojis were grouped into 4 sets according to the expressed emotion, which was identified by manually analyzing contexts:

- Joy:
  😛, 💗, 😊, 🤤, 💚, 🤗, 💘, 💙, 💋, 💓, 😂, 😇, 💝, 💜, 💖, 🧡, 💛, 🖤, ❤️
- Sadness: 😩, 🙁, 😕, 😒, 😢, 😪, 😥, 🙄, 😔, 😭, 😌, ☹️, 😓, 😰
- Anger: 😠, 😡, 👺, 🤢, 🤦‍♀️, 😐, 😬
- Uncertainty: 👀, 🧐, 🤔, 🙏, 🤞
- Neutral messages were annotated manually because they do not explicitly express emotions and include emoji quite rarely

In the next step, all messages containing non-standard (irony, sarcasm) use of emoji were excluded. The number of messages with emoji used in the direct meaning was 4500 out of 11287 messages[7] containing emoji. Whether the use of emoji was direct or non-standard was estimated manually as well. The messages from the final dataset were divided into two sets, training, and validation, containing 2300 and 2200 messages, respectively. Then, in order to obtain five binary classifiers, one for each emotion, either Logistic Regression or Multinomial Naïve Bayes was used. For this purpose, we labeled the expected class as 1 and other emotions as 0 (1-vs-all strategy). The classifiers had F1-score of 90–93%. Finally, after the classifiers predicted emotions for the rest of the text messages, only those messages were selected which were chosen by one classifier only. The resulting dataset[8] of labeled messages contains 19.000–24,000 examples of each emotion and 110.000 items in total.

4 Model Development

The proposed model was obtained with the following steps: feature extraction, model selection, new feature engineering and selection, hyperparameter tuning.

In order to extract features from the labeled data, the *sklearn TfidfVectorizer*[9] was used, yielding 29714 features. After training some traditional machine learning models, the following results were obtained (Table 2).

It should be noted that we used a validation set and the metrics used were normalized for class imbalance. For further work, we chose the logistic regression as it proved to be the best fit for the data and features that we used.

For the purpose of enhancing the model's performance, we experimented with such features as the Trager coefficient, action certainty coefficient, aggressiveness coefficient (see Sect. 2), message length in words, use of uppercase words, number of exclamation

---

[7] https://github.com/asbabiy/AIST/blob/master/msg_with_emoji.csv.

[8] https://github.com/asbabiy/AIST/blob/master/train_data.csv.

[9] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html.

Table 2. Traditional machine learning models comparison

	Precision	Recall	F1-score
Logistic regression	0.72	0.69	0.70
Naïve bayes	0.73	0.68	0.69
Random forest	0.67	0.65	0.65

marks, number of question marks and number of digits. However, we were unable to improve the F1-score, so the model's performance remained at 0.70 at this stage.

To combat overfitting, the logistic regression function was regularized. The optimal parameters were calibrated using *sklearn GridSearchCV*[10]. The final parameters of the model were: $C = 20$; $solver =$ 'lbfgs'; $max\_iter = 100000$. As a result, classifier performance (F1-score) reached 0.718 (Table 3).

Table 3. Logistic regression performance

	Precision	Recall	F1-score
Joy	0.90	0.80	0.85
Sadness	0.72	0.59	0.65
Anger	0.50	0.72	0.59
Interest	0.59	0.71	0.64
Neutrality	0.48	0.59	0.53
Weighted average	0.74	0.71	0.718

5 Discussion of Results

In order to interpret the results of the present study, the following steps were taken: 1) comparison with human performance, 2) feature importance analysis, 3) error analysis, 4) search for more features.

5.1 Human Performance

To assess the model against human performance, an experiment was conducted. Four people aged 18–19 years were asked to label a dataset containing 250 random messages from the validation set. As a result, it was found that human annotators perform only slightly better than our model: 0.74 F1-score, on average. In addition, agreement between different annotators is 70.1%. A similar result was obtained by another group of researchers [13]. Thus, we can assume that the proposed classifier has near-human performance.

[10] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

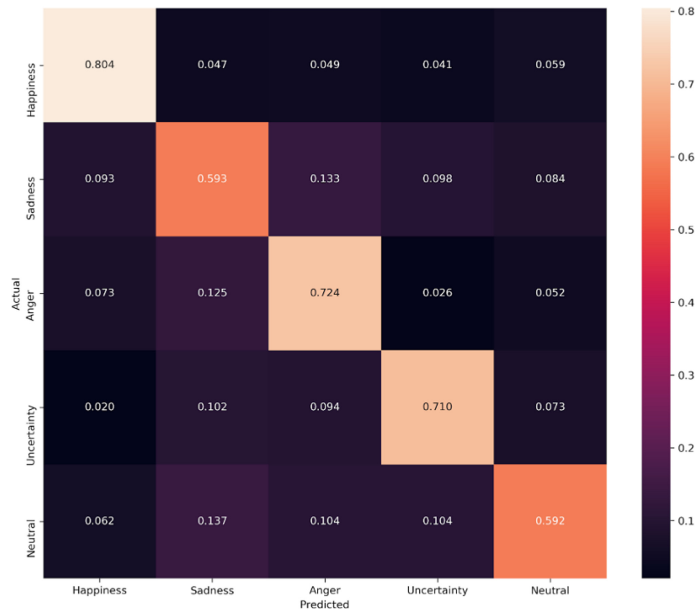

Fig. 1. Confusion matrix for emotion classes

5.2 Feature Importance Analysis

As expected, the classifier performance was different for different emotions (Fig. 1). Indeed, percentage of correctly labeled emotions ranged from 0.592 to 0.804; the Happiness class presented the least challenge, while Sadness and Neutrality were both quite difficult for the classifier. In order to get insights into the reasons for such a difference in performance, feature importance analysis was conducted.

29714 word features were extracted from the training data by TfidfVectorizer. The most important features were chosen using chi-square applied to the TF-IDF feature matrix (with respect to each class). Below are some of the lexical markers estimated to be the most important by the model:

- Happiness: молодец (well done), ура (yay), солнышко (sweetheart), благодарить (to thank), обожать (to adore), любить (to love), добрый (kind), думать (to think), удача (fortune), рад (glad).
- Sadness: скучать (to long for, to feel bored), испортить (to ruin), никогда (never), болеть (to be sick, to hurt), грустный (sad), ничего (nothing), обидеться (to get offended), бедный (poor), жаль (too bad), прощать (to forgive), плакать (to cry), завидовать (to envy), никто (nobody).
- Anger: дебилка (d*mbass), презирать (to despise), вылететь (to get kicked out, to fly out), ненавидеть (to hate), жесть (brutal, rough, creepy), жирно (greasy, brutal), стремный (weird), тупой (stupid), фу (ew), ух (ooh, ugh).
- Uncertainty: *мб* (short of *may be*), *хм* (*erm*), *почему* (*why*), *казаться* (*to seem*), *думать* (*to think*), *сколько* (*how many*, *how much*), *интересно* (*interesting*),

разве (interrogative particle), *ли* (interrogative particle), *вроде* (particle expressing uncertainty, *close to I think*, *probably*).
- Neutrality: странный (strange, weird), искренний (sincere), уйти (to go away), чисто (purely, simply), география (geography), который (which), знакомый (familiar), дз (short of homework), подняться (to go up), отличаться (to be different).

A few observations can be made. Firstly, there is a drastic difference between Uncertainty and Neutrality on the one hand and the rest of the categories on the other hand in terms of register used. Messages marked as uncertain or neutral rarely contain swear words or explicit language. Anger, in contrast, is characterized by a lot of taboo and offensive word markers. While Uncertainty and Neutrality tend to contain more sophisticated vocabulary (lower-frequency and/or more academic words), this is not true for all other classes. Furthermore, it can be noticed that negative pronouns may be associated with Sadness, while indefinite and interrogative pronouns are present in great numbers in the Uncertainty category. Finally, a lot of verbs seem to serve as Anger markers only in the imperative form. Thus, the grammatical form itself may be an important feature.

However, it is evident that these clusters of selected features are not exactly equal concerning the extent to which they describe the corresponding category. For example, the list of lexical markers for Happiness appears to be more descriptive of the target class than that of Neutrality. This naturally results in Neutrality being less recognizable by our model than Happiness. Thus, in order to provide comprehensive evaluation of each of the groups of features, it was decided to manually analyze the morphological composition and estimate the proportions of the following features in the top 50 correlating word features for each group:

- Denotative emotives - linguistic markers which contain emotional meaning in the very semantic core of it and much more often than not actualize such meaning in a speech (e.g. *ура*).
- Facultative emotives - linguistic markers which contain emotional meaning on the periphery of its semantic structure (usually as connotation) and may or may not actualize such meaning in a speech (e.g. *странный*).
- Potential emotives - linguistic markers which do not contain emotional meaning in its semantic structure and occasionally actualize such meaning in a speech, for example, as a result of secondary nomination. To a certain extent, any word may be qualified as a potential emotive as long as there might exist a communicative situation in which this word would obtain an emotional connotation.

Concerning potential emotives, it needs to be noted that features which fall into this category are hardly of any use in terms of helping to automate emotion identification. That is because they are not universal emotion markers but occasional. Thus, the fact that they appear in the list of the most correlated features may only mean two things: firstly, for people from our current sample this word does carry some emotional meaning, secondly, there is not enough stronger features (denotative or, at least, facultative ones) for this exact category in this exact dataset. Consequently, there is little point in considering these features in further development as we are aiming at building a classifier that would work for messages from a bigger variety of authors than just from those who are represented in the current dataset.

Moreover, a scale of expressiveness was also added, indicating to what extent the words in a group accentuate the corresponding emotion.

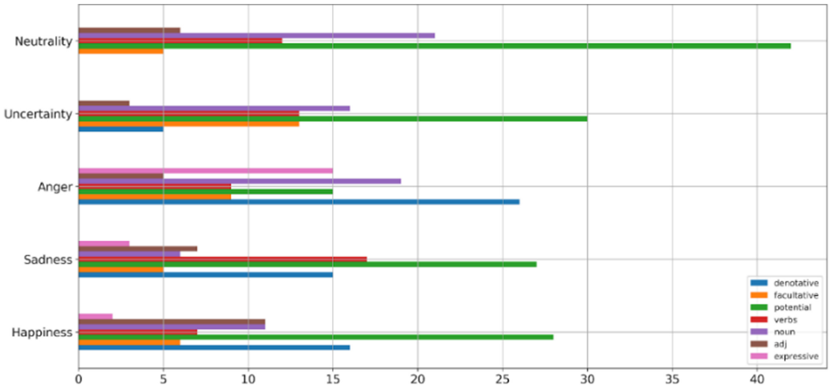

Fig. 2. Feature analysis

It is quite obvious from the graph (Fig. 2) that the denotative emotive rate and the expressiveness rate appear to be the most fluctuating values while the facultative emotive rate seems to be the most stable one among all the categories.

Since there are certain differences in recognition rates for each of the categories (as shown in the confusion matrix above, Fig. 1), it seems logical to conclude that such features as denotative emotives, adjectives and expressiveness are crucial for emotion recognition, as they are the most highly-correlated: the Pearson coefficient values are 0.727, 0.816 and 0.321, respectively. Thus, the more denotative emotives are found for a category, the greater is the probability of its recognition by the model.

5.3 Error Analysis

While Happiness is found to be the most recognizable category, Sadness and Neutrality pose the most challenge. Anger and Uncertainty are placed in between. Figure 3 demonstrates how often the two categories were confused for each other.

Statistics indicate that top three error types are Sadness being confused with other non-positive emotions. Regarding Sadness and Anger confusion, a number of messages in which sad feelings are conveyed contain a lot of expressive words such as swear words, perceived by the model as a feature of the Anger class. Thus, 13.3% Sadness messages were classified as Anger, e.g. *я послан нах*й* (*I'm told to f*ck off*), *с*кааа плачу… первый трек без тома* (*b*tch I'm crying… first track without Tom…*). On the other hand, not all aggressive messages contain explicit language. Hence, they are quite likely to be misclassified as Sadness. For example, *не бейте карину по ноге* (*don't hit Karina on the leg*), *света не общайся со мной больше* (*Sveta don't talk to*

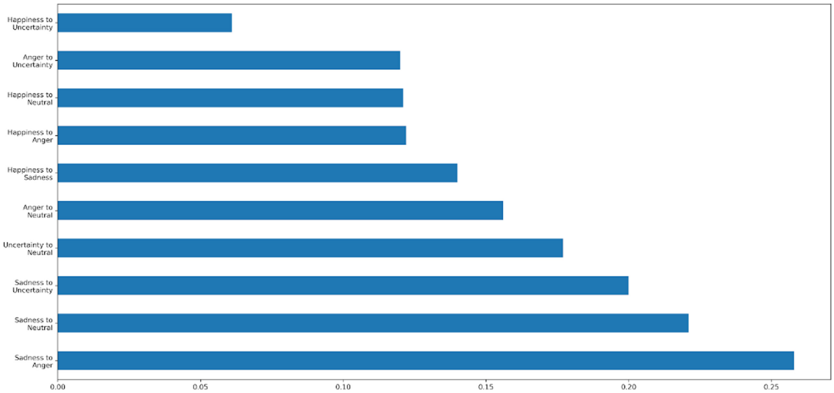

Fig. 3. Pairwise confusion of categories

me anymore), *котьку не обижай* (*don't hurt the kitten*). This error occured in 12.5% of cases with Anger messages.

As for the lack of distinction between Sadness and Neutrality, it is, perhaps, due to the total absence of denotative emotives in the Neutrality class. Consequently, whenever a neutral message contains a potential emotive of Sadness, it is misclassified as such. For instance, *хотя она ж не обижалась* (*though she didn't resent*), *я бы не ставил какие то философские вопросы в проблему* (*I wouldn't put some philosophical issues to the problem*), *ну а что не так* (*so what is wrong*). Such cases amount to 13.7% of all Neutrality items. However, 8.4% of Sadness messages were labeled as Neutrality, too. Since it is noticeable that Neutral markers are plainly random and do not reflect the specifics of a neutral emotional state, it is believed that Sadness may be misclassified as Neutrality only under the following circumstances: firstly, the sad message must have little to no emotive markers of Sadness found by the model, and secondly, there must be several markers of Neutrality. To illustrate, there are a few actual examples: *Не смогли бы записаться и уйти* (*Wouldn't be able to sign up and go*), *Он был не искренним* (*He wasn't sincere*).

Speaking of the reason for which Sadness and Uncertainty tend to be confused, one interpretation is that such emotions are often present in a message simultaneously since it is natural for a person to be upset and uncertain at the same time. Sadness was misclassified as Uncertainty in 9.8% cases (for example, *мне 14 что я вообще здесь делаю* (*I'm 14 what am I even doing here*), *как все сложно* (*everything is so complicated*), while Uncertainty was identified as Sadness in 10.2% cases (e.g. *но мучают сомнения поедем ли* (*but I'm confused if we'll go*), *я хз как это работает* (*I dunno how it works*)).

Taking into consideration all of the above, it is expected that the following steps will help eliminate errors and enhance the model's performance:

- correct the model's behavior concerning explicit language,
- exclude random word features from Neutrality and manually find and add more distinctive markers (not only words, but probably other linguistic levels as well),

- manually add a greater number of denotative emotives to the list of features,
- use a different emotion classification and/or a way to account for multiple emotions expressed in one message.

5.4 Search for More Features

For the purpose of finding more features, it was decided to test the new markers obtained during feature and error analysis, using tree-based ensemble models. There are 27 markers in total. These markers include:

- Morphological features of messages, such as the presence, as well as the absolute and relative frequencies of numerals, adjectives, nouns, verbs (especially imperatives), function words, and negative, indefinite, and interrogative pronouns.
- Formal and lexical markers such as the maximal and average word length in a message, the presence of special words of etiquette like *спасибо*, *пожалуйста* or *простите*, and so forth.
- Graphic/punctuation indicators of expressiveness such as uppercase words, emojis, exclamation and question marks.
- The presence and proportion of explicit language.

Tree-based ensemble models used for detecting significant features include Extra-Trees Classifier, Light Gradient Boosting Machine, and AdaBoost Classifier. By means of *feature_importances_* attribute, the following results were obtained (Fig. 4).

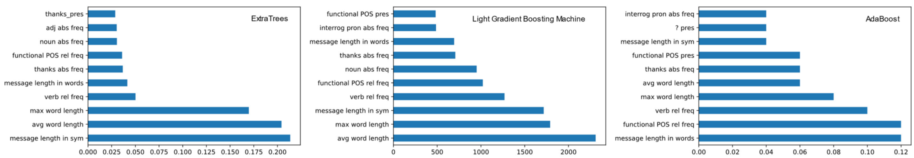

Fig. 4. Feature importance according to tree-based ensemble models

As it is evident from the presented graphs showing top ten important features, the following markers were selected as more significant by all three models: noun count, *спасибо* count, ratio of function words, message length in words, verb ratio, maximal length of a word in the message.

The presence of function words, the number of interrogative pronouns, and message length in symbols seem less significant as they managed to reach top ten of just two tree-based ensemble models out of three. Finally, the presence of *спасибо* and the number of adjectives were also found important by the ExtraTrees model. Thus, this makes a list of features worth implementing in future work.

Apart from this, it is also necessary to work out a way of distinguishing the register of words in a message as it might improve the model's performance by enhancing the recognizability of Neutrality and Uncertainty.

6 Conclusion

In this paper, we addressed the problem of developing an integrated approach to emotion classification combining linguistic analysis and machine learning. We propose an automatic classifier of emotions present in informal text messages in Russian. We also present a new dataset that was based on processing 1800000 messages. A labeled training set was generated semi-automatically: five binary classifiers (one for each emotion) were trained on a smaller manually annotated sample of 4500 messages. The performance of the proposed final classifier of emotions is 0.718 (F1-score). It is estimated to be close to human performance on this task.

Feature importance analysis has shown that denotative emotives, level of expressiveness, morphological composition and stylistic markers play a significant role. However, it is difficult to distinguish between non-positive emotions, as shown in our error analysis. Future work includes reducing the model's mistakes by changing its behavior concerning explicit language, manually excluding insignificant features, and adding important ones such as function words ratio, maximal word length in a message, verb ratio, etc. In addition, it is believed that further study will benefit from feature importance information obtained by using the eli5[11] package. Moreover, we plan to expand the emotion classification by adding the emotions of contempt and disgust and differentiating between fear and surprise. Finally, we plan to implement multi-label classification, as there are cases when the same short text or even a single sentence contains more than one emotion.

Acknowledgments. The work is supported by RSF (Russian Science Foundation) grant 20-71-10010.

References

1. Abdul-Mageed, M., Ungar, L.: Emonet: Fine-grained emotion detection with gated recurrent neural networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1: Long papers, pp. 718–728 (2017)
2. Canales, L., Strapparava, C., Boldrini, E., Martnez-Barco, P.: Exploiting a bootstrapping approach for automatic annotation of emotions in texts. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 726–734. IEEE (2016)
3. Chatterjee, A., Narahari, K.N., Joshi, M., Agrawal, P.: Semeval-2019 task 3: emocontext contextual emotion detection in text. In: Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 39–48 (2019)
4. Ekman, P.: Expression and the nature of emotion. Approaches Emot. **3**(19), 344 (1984)
5. Gaind, B., Syal, V., Padgalwar, S.: Emotion detection and analysis on social media (2019). arXiv preprint arXiv:1901.08458.
6. Gudovskikh, D.V., Moloshnikov, I.A., Rybka, R.B.: Analiz emotivnosti tekstov na osnove psikholingvisticheskikh markerov s opredeleniem morfologicheskikh svoystv. Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya (3) (2015)

[11] https://eli5.readthedocs.io/en/latest/.

7. Gupta, U., Chatterjee, A., Srikanth, R., Agrawal, P.: A sentiment-and-semantics-based approach for emotion detection in textual conversations (2017). arXiv preprint arXiv:1707.06996.
8. Hasan, M., Rundensteiner, E., Agu, E.: Automatic emotion detection in text streams by analyzing twitter data. Int. J. Data Sci. Anal. **7**(1), 35–51 (2019)
9. Klinger, R.: An analysis of annotated corpora for emotion classification in text. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 2104–2119 (2018)
10. Babii, A., Kazyulina, M., Malafeev. A.: Automatic emotion identification in russian text messages. In: Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”, no. 19, Supplementary volume, pp. 1002–1010 (2020)
11. Mehendale, N.: Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. **2**(3), 1–8 (2020)
12. Mohammad, S.M., Kiritchenko, S.: Using hashtags to capture fine emotion categories from tweets. Comput. Intell. **31**(2), 301–326 (2015)
13. Mohammad, S.M., Zhu, X., Kiritchenko, S., Martin, J.: Sentiment, emotion, purpose, and style in electoral tweets. Inf. Process. Manag. **51**(4), 480–499 (2015)
14. Pazel’skaya, A.G., Solov’ev, A.N.: Metod opredeleniya emotsiy v tekstakh na russkom yazyke. In Komp’yuternaya lingvistika i intellektual’nye tekhnologii: Po materialam yezhegodnoy Mezhdunarodnoy konferentsii «Dialog» (Bekasovo, 25–29 maya 2011 g.). M.: Izd-vo RGGU, no. 10, p. 17 (2011)
15. Seyeditabari, A., Tabari, N., Gholizadeh, S., Zadrozny, W.: Emotion detection in text: focusing on latent representation (2019). arXiv preprint arXiv:1907.09369.
16. Zahiri, S.M., Choi, J.D.: Emotion detection on tv show transcripts with sequence-based convolutional neural networks. In: Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence (2018)