AUTOMATIC EMOTION IDENTIFICATION IN RUSSIAN TEXT MESSAGES

Babii A. S. (bac03704@gmail.com),
Kazyulina M. S. (mskazyulina@edu.hse.ru)
National Research University Higher School of Economics
(Russia, Nizhny Novgorod)

Malafeev A. Yu.

Research advisor

Automatic emotive text analysis has demonstrated its relevance in recent years. In this paper, we address the issue of identification emotions in the text of informal internet-discourse of the Russian language. We consider text messages collected from Telegram and VK. Due to difficulty of such advanced form of sentiment analysis, this paper proposes an integrated approach to combining linguistic methods and machine learning. As a result, an automatic classifier of text messages on expressed emotions is designed. On testing, our model is estimated to provide near-human performance.

Keywords: machine learning, emotion identification, emotiveness, sentiment analysis, natural language processing

1. Introduction

Nowadays, NLP is one of the fastest-growing areas. Achievements in the field of automatic natural language processing include technologies that help accurately identify the sentiment of the text. These technologies are usually based on either linguistic text analysis (a rule-based approach) or machine learning.

Thus, the present research focuses on the development of an integrated approach that combines linguistic analysis and machine learning for the Russian language. Moreover, though current research it is intended to design an automatic classifier of informal internet-discourse text messages with explicit emotions in them.

The relevance of the present study is determined by the following aspects:

1) A high potential for practical use for automatic emotion recognition in a text.
2) The need for systematization of knowledge on linguistic data application in the field of NLP.
3) The need for an integrated approach (combining linguistic data and machine learning) for the design of automatic semantic analysis models.

1.1. Emotiveness

Emotiveness as an immanent function of a language is carried out by the aid of emotive features on different language levels. Several studies have found that lexical emotive features appear effective in terms of automatic text analysis. Usually, such studies are conducted by the virtue of dictionaries containing expressive words [Pazelskaya, Solovyov, 2011].

Other than that, morphological emotives were successfully used in the development of a system for predicting the emotiveness of texts [Gudovskikh, Moloshnikov, Rybka, 2015]. These emotive markers are the Trager coefficient, the coefficient of certainty of action, and also the coefficient of aggressiveness.

Furthermore, prosody appears to be the most expressive level of language as by the aid of intonation any of the linguistic means may be actualized in speech as an emotive feature [Mairesse, Polifroni, Fabbrizio, 2012]. Despite the fact that there is a significant difficulty in identifying intonation of the written words, there are certain indicators such as exclamation marks, question marks, and uppercase words. Taking all the aforementioned features into consideration may help indicate expressed emotions in the text.

1.2. Machine learning

Machine learning algorithms are proven to be effective in terms of detection, classification, and quantification of emotions of text in any form [Gaind, Syal, Padgalwar, 2019]. However, shortcomings of classical methods of machine learning include the fact that they do not allow to perform sentiment analysis on a linguistic level that is deep enough. Nevertheless, it is supposed that these limitations might be compensated for by implementing linguistics emotiveness analysis [Prabowo et al., 2009]; [Konig, 2006].

1.3. Emotion classification

With the purpose of developing an automatic classifier of text messages on expressed emotions, it is required to adopt a classification of emotions. As of now, there is no generally accepted emotion classification. Yet there is P. Ekman classification [Ekman, 1984] which appears the most popular one. It includes happiness, surprise, sadness, anger, fear, disgust, and contempt.

However, in order to enhance the effectiveness of the classifier, the categories of fear and surprise were merged into one category of uncertainty, and the categories

of disgust and contempt were removed due to lack of appropriate data. Thus, the classifier developed in this paper divides messages into five groups: happiness, sadness, anger, uncertainty, and neutrality. The last category includes texts without explicit emotions.

2. Data

The present study observes informal internet-discourse of personal messaging. The creation process of message corpus was divided into 4 stages:
- sourcing and parsing data;
- data preprocessing;
- data processing;
- corpus annotation.

2.1. Parsing and processing data

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 messages written by the user but also the ones they received. This allows us to request access and use personal messages and public chats from VKontakte[1] and Telegram[2] to create our message corpus. We chose these social networks as they are the most popular ones with the Russian-speaking population. Owing to this fact, dataset containing messages from private and public VK and Telegram chats is supposed to be representative as it is thematically diverse and stylistically even.

Regarding parsed data, statistical population includes people that are 12–35 y.o. with educational levels from general to postgraduate. The statistical population is represented by a sample of 4,584 people.

Statistically, the average length of parsed messages is 4–5 words. In the process of normalization, all the messages consisting of more than 9 words were removed, and so were tags, hashtags, and messages with no Cyrillic symbols. Furthermore, e-mail addresses and phone numbers were coerced to tokens <email> and <phone>, respectively. Then the corpus was normalized with regular expressions and lemmatized by *rnnmorph*[3] lemmatizer. The final dataset of lemmatized messages consisted of 1,800,000 messages.

---

[1]   https://vk.com

[2]   https://telegram.org

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

2.2. Corpus annotation

In order to annotate the messages with emotion labels correctly, it was decided to use emoji ideograms as an objective emotion indicator. Obviously, not all the messages in the corpus contain emojis. Thus, it was decided to build 5 classifiers (one for each emotion) trained on those objectively labeled messages. Those classifiers then labeled the rest of the data. All the details concerning annotation process are below.

Firstly, the emojis used less than 40 times overall were removed from the observed data. The rest of the emojis formed 4 sets grouped by expressed emotion which was identified through context analysis. It was performed manually with excluding ambiguous cases from the dataset.

- Happiness: 😙, 💗, 😊, 😌, 💚, 😋, 💘, 💙, 🥖, 💜, 😂, 😊, 💝, 💜, 💓, 🧡, 💛, 🖤
- Sadness: 😩, 🙁, 😔, 😒, 😟, 😪, 🥺, 😐, 😞, 😓, 😕, 😖, 😰
- Anger: 😠, 😡, 🤬, 😤, 🤦, 😑, ⌨
- Uncertainty: 👀, 🥴, 🤨, 🙏, 🤞
- Neutral messages were selected manually from messages without emojis because samples of this class do not express emotions and include emoji quite rarely.

Secondly, all the messages containing the occasional use of emoji were excluded manually. The dataset of neutral messages and messages with emoji used in the plain meaning amounted to approximately 4,500 samples. Messages from the final dataset were divided into 2 sets: an initial training set and validation set, containing 2,300 and 2,200 messages, respectively.

Thereafter, in order to design 5 binary classifiers for each emotion either Logistic Regression or Multinomial Naïve Bayes was used. For this purpose, we labeled the expected class as "1", randomly mixed messages which were labeled as other emotions and put them into class "0". To extract features from the labeled data, TF-IDF was used. It led to model performance reaching 90–93% for all these classifiers.

Finally, only those messages marked only once with "1" were selected. The resulting dataset of the labeled messages contains ~22,000 examples of each emotion and 110,000 items in total.

3. Development of a model for multiclass classification of emotions

Development of a multiclass classifier of text messages for the expressed emotion requires the following steps:

1) feature extraction;
2) model choice;
3) feature engineering and feature selection;
4) hyperparameter tuning.

3.1. Feature extraction and model choice

In order to extract features from the labeled data, *TfidfVectorizer*[4] was used. As a result, 29,714 features were revealed. After training statistical models, the following results were obtained. For further work, we chose the logistic regression as it proved to be the best for such classification task.

Table 1

	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

3.2. Feature engineering and feature selection

For the purpose of enhancing model performance accuracy, the features below were constructed and analyzed:

- lexical denotative emotive features;
- Trager coefficient;
- action certainty coefficient;
- aggressiveness coefficient;
- message length in words;
- use of uppercase words;
- number of exclamation marks;
- number of question marks;
- number of digits.

None of those features appeared relevant for the collected data except for the following:

1) Number of digits.
   According to the graph in **Fig. 1**, Neutral class is more prevalent among all the messages containing digits.

2) Number of exclamation marks.
   It is obvious from **Fig. 2** that Neutral class includes the fewest number of exclamation marks overall which may also serve as a marker for identifying emotion.

3) Use of uppercase words.

---

[4] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
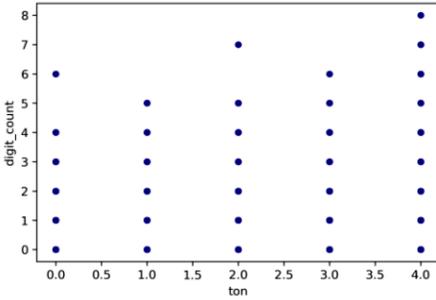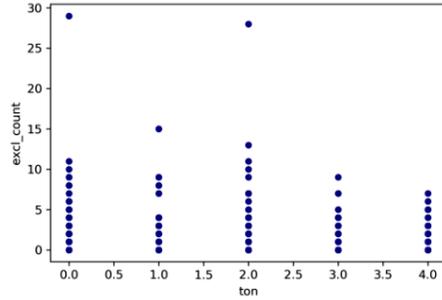
Fig. 1

Fig. 2

There are several factors that lead to inapplicability of the other constructed features. Concerning morphological emotive markers, it is supposed that the analyzed texts are not long enough to have an emotively representative grammar structure. As for the rest, they were distributed almost evenly among all the classes and thus did not seem to be effective in recognizing the emotion expressed in the text. It is possible that the sample used in this study was not representative enough to reflect some of the ways emotions may be conveyed via text. This implies that training statistic model on a different dataset may lead to different results concerning the relevance of these emotive features.

However, retraining the logistic regression model with such features as a number of digits, a number of exclamation marks and use of uppercase words did not help to increase the accuracy, making the data noisier. Thereby, after feature engineering and feature selection, model F1-Score remained 0.70.

3.3. Hyperparameter tuning

With a view of reducing overfitting, the function was regularized. The optimal parameters were calibrated using *GridSearchCV*. The final parameters of the model: C = 20; solver = 'lbfgs'; max_iter = 100,000. As a result, F1-Score increased up to 0.718.

Table 2

	Precision	Recall	F1-Score
Happiness	0.90	0.80	0.85
Sadness	0.72	0.59	0.65
Anger	0.50	0.72	0.59
Uncertainty	0.59	0.71	0.64
Neutral	0.48	0.59	0.53
Weighted average	**0.74**	**0.71**	**0.718**

3.4. Error analysis and interpretation

Error analysis has shown that Happiness was recognized more often, while Neutral was correctly identified less frequently. As shown by chi-squared test, messages with explicit Happiness in almost all cases contain a great number of specific emotive markers which makes it relatively easy to identify this emotion in text. On the contrary, messages with no expressed emotion (class Neutral) contain the least wide range of distinctive features. This leads to difficulty in finding a correlation between certain characteristics of the text and the label.

Furthermore, it is noticeable from this confusion matrix that non-positive messages were quite often misclassified as Sadness. Such messages often contain the following uni- and bigrams: "блин", "как бы", "не", "по вам", "никто не", "хз". It is obvious from this list that misclassified messages contain lexical markers that may express not just one, but several emotions. It is due to human ability to experience a cluster of emotions.

Sadness was often misclassified as Anger. Such messages usually contain the following words: "убить", "ой", "ещё бы", "за что", "неужели" and some foul words. Owing to the fact that class Sadness has relatively few distinctive markers, while class Anger possesses a lot of quite strong ones, it is reckoned that whenever a message with explicit emotion of sadness contains an Anger marker, it may be easily misclassified. The process of elimination of this type of mistakes is supposed to include adding more Sadness markers manually.

To assess the model's performance, an experiment involving human perception of emotions from the text was conducted. Four people aged 18–19 years were asked to label a dataset containing 250 random messages from a validation set. As a result, it was found that people embark on the task of message classification on expressed emotion with the average accuracy of 74%. In addition, consent between different people is ~70%. However, it should be stated that human perception of emotions expressed in text depends on one's emotive competence and will not be the same for all people. Thus, it is required to conduct a larger experiment. Nonetheless, as of now it is supposed that this classifier has a near-human performance.

4. Conclusion and Future Work

In current paper, the problem of the development of an integrated approach that combines linguistic analysis and machine learning was addressed. Through present research an automatic classifier of informal internet-discourse text messages in Russian with explicit emotions in them was designed. In order to do this, 1,800,000 messages were collected and processed. From those messages the dataset of 110,000 samples was formed and used to train and test logit model. A labeled training set was generated almost automatically: five binary classifiers (one for each emotion) were designed. As a result, the accuracy of the developed automatic classifier of informal internet-discourse text is 71.8%. It is estimated to be close to human performance on this task. However, well-designed experiment is yet to be conducted.

Error analysis has shown that there is difficulty in distinguishing between nonpositive emotions. In order to eliminate errors of this kind, it is suggested to implement a multilabel classification instead of the multiclass one. Furthermore, future work includes correcting such mistakes by implementing a larger number of emotive markers for Sadness, Uncertainty and Neutral. Moreover, it is envisaged to expand emotion classification by adding contempt and disgust and differentiating fear and surprise. In addition, it is believed that present study will benefit from the information concerning feature importance which will be found by using eli5[5] package. Other than that, in future, it is intended to implement state-of-the-art technologies to improve classifier performance.

References

1. *Ekman, P.* (1984). Expression and the nature of emotion. Approaches to emotion, 3(19), 344.
2. *Gaind, B., Syal, V., & Padgalwar, S.* (2019). Emotion detection and analysis on social media. arXiv preprint arXiv:1901.08458.
3. *Gudovskikh, D. V., Moloshnikov, I. A., Rybka, R. B.* (2015). Analiz emotivnosti tekstov na osnove psikholingvisticheskikh markerov s opredeleniem morfologicheskikh svoystv. Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya, (3).

---

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

4. *Konig A. C., Brill E.* (2006) Reducing the Human Overhead in Text Categoriza-
    tion. Proceedings of the12th ACM SIGKDD Conference on Knowledge Discovery
    and Data Mining: 598–603.
5. *Mairesse, F., Polifroni, J., Di Fabbrizio, G.* (2012, March). Can prosody inform
    sentiment analysis? experiments on short spoken reviews. In 2012 IEEE In-
    ternational Conference on Acoustics, Speech and Signal Processing (ICASSP),
    pp. 5093–5096. IEEE.
6. *Pazel'skaya A. G., Solov'ev A. N.* (2011, May). Metod opredeleniya emotsiy v tek-
    stakh na russkom yazyke. In Komp'yuternaya lingvistika i intellektual'nye tekh-
    nologii: Po materialam yezhegodnoy Mezhdunarodnoy konferentsii «Dialog»
    (Bekasovo, 25–29 maya 2011 g.). M.: Izd-vo RGGU (No. 10, p. 17).
7. *Prabowo R., Thelwall M.* (2009) Sentiment Analysis: A Combined Approach.
    Journal of Informetrics, 3(2).