DOI: 10.15514/ISPRAS-2021-33(3)-14

Language Models Application in Sentiment Attitude Extraction Task

[1]*Rusnachenko N. L., ORCID: 0000-0002-9750-5499 <rusnicolay@gmail.ru>*
[1] *Bauman Moscow State Technical University,*
5, Building 1, 2-nd Baumanskaya Str., Moscow, 105005, Russia

Abstract. Large text can convey various forms of sentiment information including the author's position, positive or negative effects of some events, attitides of mentioned entities towards to each other. In this paper, we experiment with BERT based language models for extracting sentiment attitudes between named entities. Given a mass media article and list of mentioned named entities, the task is to extract positive or negative attitudes between them. Efficiency of language model methods depends on the amount of training data. To enrich training data, we adopt distant supervision method, which provide automatic annotation of unlabeled texts using an additional lexical resource. The proposed approach is subdivided into two stages FRAME-BASED: (1) sentiment pairs list completion (PAIR-BASED), (2) document annotations using PAIR-BASED and FRAME-BASED factors. Being applied towards a large news collection, the methods generates RuAttitudes2017 automatically annotated collection. We evaluate the approach on RuSentRel-1.0, consisted of mass media articles written in Russian. Adopting RuAttitudes2017 in the training process results in 10-13% quality improvement by F1-measure over supervised learning and by 25% over the top neural network based model results.

Keywords: Sentiment Analysis, Relation Extraction, Distant Supervision, Neural Networks, Language Models

For citation: Rusnachenko N. L. Language Models Application in Sentiment Attitude Extraction Task. Trudy ISP RAN/Proc. ISP RAS, 2021, vol. 1, issue 3, pp. 199–222. 10.15514/ISPRAS-2021-33(3)-14

Acknowledgments: This work was supported by a grant from the RFBR 20-07-01059

Analysis of the influence of pre-training on the distribution of weights of the attention mechanism in language models

To analyze the contribution of various elements of the context to the result obtained, the weights of the attention mechanism are often compared. The following states of language models were chosen for analysis: mBERT, SentRuBERT and SentRuBERT-NLI$_P$ (pre-trained version of SentRuBERT by the RuAttitudes2017$_{\text{Large}}$ collection). Among the entire set of contexts, only those contexts are considered that were extracted by the pretrained model SentRuBERT (NLI$_P$ + NLI) from the test set of the collection RuSentRel. Thus, 1032 contexts were analyzed. In the contexts, the occurrences of the lexicon of evaluative words of the Russian language RuSentiLex [1] (SENTIMENT) and the occurrences of frames (FRAMES) are additionally marked in the contexts.

Таблица 1: The average estimate of the probability of attention over the heads of the BERT language model for each of the 12 layers separately for: class tokens (CLS), separators (SEP), participants in the relationship all third-party tokens to FRAMES and SENTIMENT separately; the top values in the rows are marked in bold

	Layer Number											
Terms group	1	2	3	4	5	6	7	8	9	10	11	12
mBERT												
[CLS]	0.06	0.33	**0.36**	0.29	0.31	0.06	0.04	0.04	0.05	0.06	0.07	0.04
SEP	0.04	0.07	0.06	0.06	0.07	0.09	0.09	0.11	**0.12**	0.09	0.09	0.07
$\underline{E}_{subj}/\underline{E}_{obj}$	0.05	0.04	0.04	0.06	0.04	0.06	0.06	0.06	0.06	**0.07**	**0.07**	0.05
others→FRAMES	**0.07**	0.03	0.03	0.03	0.03	0.05	0.04	0.05	0.04	0.04	0.03	0.03
others→SENTIMENT	**0.08**	0.04	0.03	0.03	0.04	0.05	0.04	0.05	0.05	0.04	0.03	0.04
SentRuBERT												
[CLS]	0.03	0.27	0.33	0.30	**0.39**	0.09	0.02	0.03	0.03	0.05	0.04	0.02
SEP	0.05	0.06	0.03	0.04	0.04	0.15	0.22	**0.39**	0.28	0.29	0.07	0.04
$\underline{E}_{subj}/\underline{E}_{obj}$	0.10	0.06	0.07	0.07	0.05	0.06	0.08	0.04	0.06	0.05	0.11	**0.12**
others→FRAMES	0.05	0.03	0.03	0.03	0.03	0.04	0.04	0.03	0.05	0.05	**0.07**	0.06
others→SENTIMENT	0.06	0.03	0.03	0.03	0.03	0.04	0.04	0.05	0.06	0.06	**0.08**	**0.08**
SentRuBERT-NLI$_P$												
[CLS]	0.03	0.27	**0.36**	0.31	0.34	0.05	0.01	0.02	0.01	0.02	0.02	0.02
SEP	0.06	0.04	0.03	0.05	0.04	0.20	0.20	**0.28**	**0.28**	**0.28**	0.04	0.08
$\underline{E}_{subj}/\underline{E}_{obj}$	0.10	0.07	0.08	0.08	0.07	0.07	0.09	0.06	0.07	0.11	**0.28**	0.23
others→FRAMES	0.07	0.04	0.04	0.04	0.05	0.06	0.05	0.07	0.07	0.05	**0.10**	0.08
others→SENTIMENT	0.08	0.05	0.05	0.04	0.05	0.07	0.06	**0.09**	0.08	0.07	0.08	**0.09**

For each input context s tokens long, the attention weight vector $a \in R^{l \times h \times s \times s}$ contains the values of each layer, for each model head BERT (l is the number of layers of the language model; h is the number of heads). For particular layer l' and a head h', the matrix $a_{l',h'} \in R^{s \times s}$ describes the link weights of the layer's input data tokens l' with its output (next layer tokens):

- [CLS] – class;
- [SEP] – sequence boundaries;
- [S/O] – relationship participants ($\underline{E}_{subj}/\underline{E}_{obj}$);
- Groups FRAMES and attention to them by other context tokens;

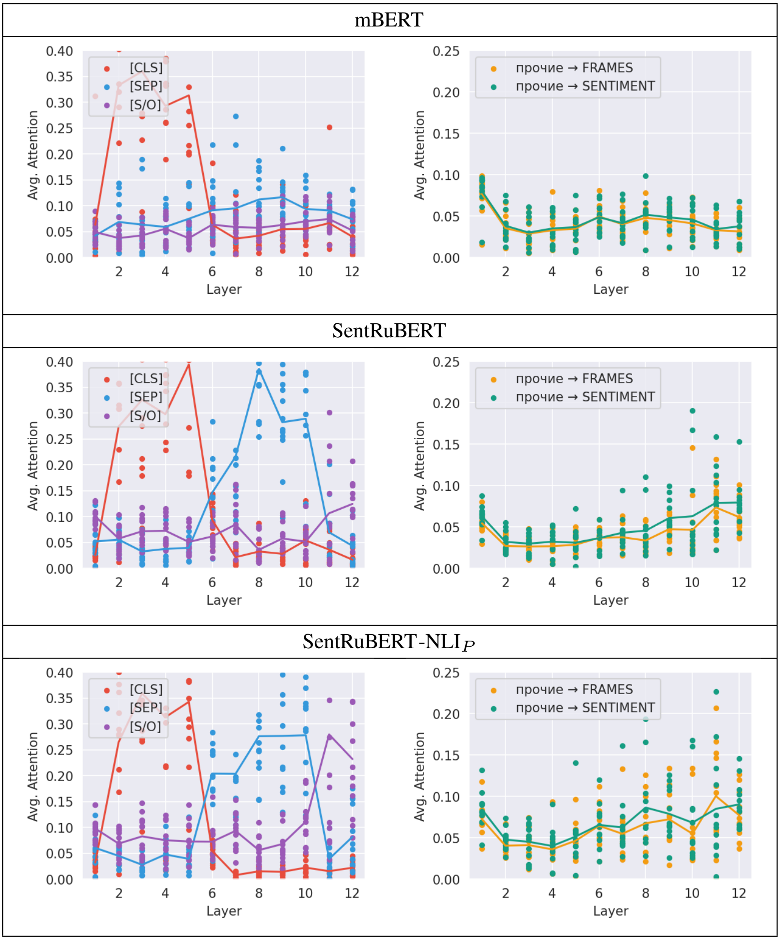

Рис. 1: Layered assessment of the distribution of attention of language models BERT to tokens [CLS], [SEP], objects and subjects of the relation [S/O] (left column) and frames and sentiment words (right column); the lines connect the average values of the weights of each layer of the model [2]

- Groups SENTIMENT and attention to them by other context tokens.

Fig. 1 illustrates the layer-by-layer evaluation of the attention weights for the given groups of tokens. Average values for each layer are shown in[1] the Table 1.

[1]For average estimates for the groups FRAMES and SENTIMENT, only such contexts are taken into account that contain at least one occurrence of the term of the corresponding group. As a result, 68% of contexts were taken into account in the «other→FRAMES» statistics, and 75% in statistics «other→SENTIMENT»

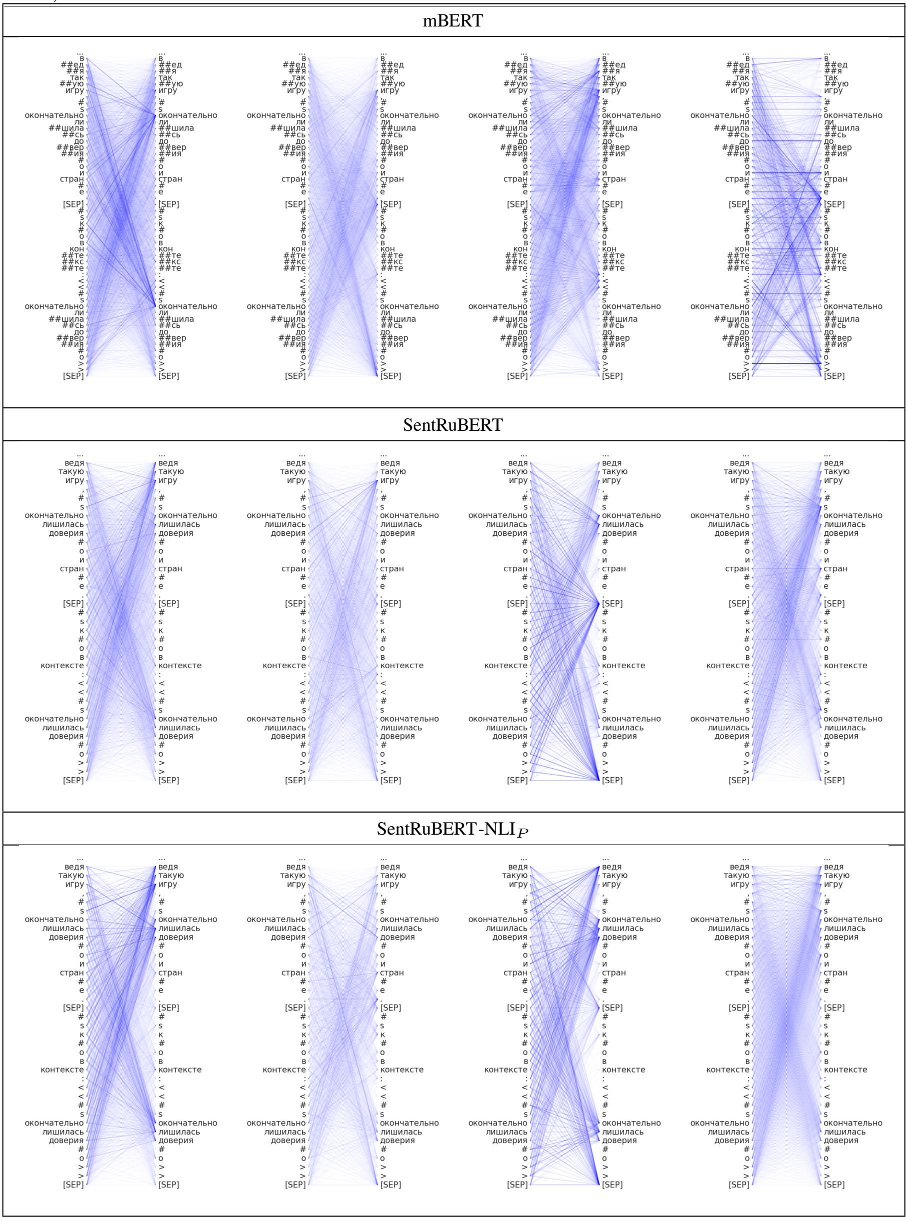

Рис. 2: An example of visualization [2] of head weights #2 (layers from left to right: 2,4,8,11) as the evolution of the attention of the mBERT model in the finetuned models SentRuBERT and SentRuBERT-NLI$_P$

It should be noted the high rates of attention to the class token [CLS] on layers 2-5 up to 35-40%. For SentRuBERT, there is an increase in attention on the tokens [SEP] (layers 7-10) and [S/O] (on the final layers). There is also an increase in attention on the tokens FRAMES and SENTIMENT from others tokens on the final layers up to 7-10%. The use of indirect learning (SentRuBERT-NLI$_P$) increased attention to [S/O] in the final layers: the weights doubled when compared with SentRuBERT. There is also an additional increase in attention to the SENTIMENT and FRAMES tokens from other tokens in the middle and end layers.

In order to visually illustrate the effect of additional training, Fig. 2 shows the visualization of head weights No. 2 for each analyzed model BERT, by layers (from left to right) 2, 4, 8, 11 of the following example:

«Ведя такую игру, $\underline{E}_{subj}$ *окончательно лишилась доверия* $\underline{E}_{obj}$ *и стран* $\underline{E}$*.* $\underline{E}_{subj}$ *к* $\underline{E}_{obj}$ *в контексте «*$\underline{E}_{subj}$ *окончательно лишилась доверия* $\underline{E}_{obj}$*».* (*While playing such a game,* $\underline{E}_{subj}$ *has completely lost the trust of* $\underline{E}_{obj}$ *and* $\underline{E}$ *countries.* $\underline{E}_{subj}$ *to* $\underline{E}_{obj}$ *in the context of «*$\underline{E}_{subj}$ *is completely distrusted by* $\underline{E}_{obj}$)

In the SentRuBERT-NLI$_P$ model, among others, the focus of attention is most pronounced on the occurrences of the «finally» and «lose trust» frames (layer 8).

References

[1]. N. Loukachevitch и A. Levchik. Creating a general russian sentiment lexicon. В *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, страницы 1171—1176, 2016.

[2]. K. Clark, U. Khandelwal, O. Levy и C. D. Manning. What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341*, 2019.