

Russian aspect-based sentiment analysis

Mamonov K. R.

`mamonov.kr@phystech.ru`

Moscow Institute of Physics and Technology

This report describes the part of work that was conducted during the course Deep Learning in Natural Language Processing. There was investigated the Topic Modelling approach as way to construct the additional features for Neural Network and implemented the ULMFit model with accuracy 0.62.

Keywords: *sentiment analysis, deep learning, aspects.*

1 Data

We use the data from competition Dialogue-2020. Several steps of processing were performed: texts were lemmatized, tokenized, all trash was dropped. There were selected hashtags, links and other tags that I used as metadata.

2 Topic Modelling

Using BIGArtm [2,3] library I constructed PLSA topic model, chose the optimal parameters of the model to make it interpreted and clustered all comments according topic that it devote to. The results is shown on Figure 1.

It is not so well as I expected but it is quiet ok as one of the features.

3 ULMFit

The next my task was implementation of ULMFit model [1]. That I have successfully implemented only for English tweets yet with accuracy 0.62, train loss 0.911, validation loss 0.916. But thanks to universality of this model it will not take much time to adapt it for Russian. Wordclouds for negative and positive sentiments you can see on Figure 2 and Figure 3.

References

- [1] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 328–339. Association for Computational Linguistics, 2018.
- [2] Konstantin Vorontsov, Oleksandr Frei, Murat Apishev, Peter Romov, and Marina Dudarenko. Bigartm: open source library for regularized multimodal topic modeling of large collections. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 370–381. Springer, 2015.
- [3] Konstantin Vorontsov and Anna Potapenko. Additive regularization of topic models. *Machine Learning*, 101(1-3):303–323, 2015.

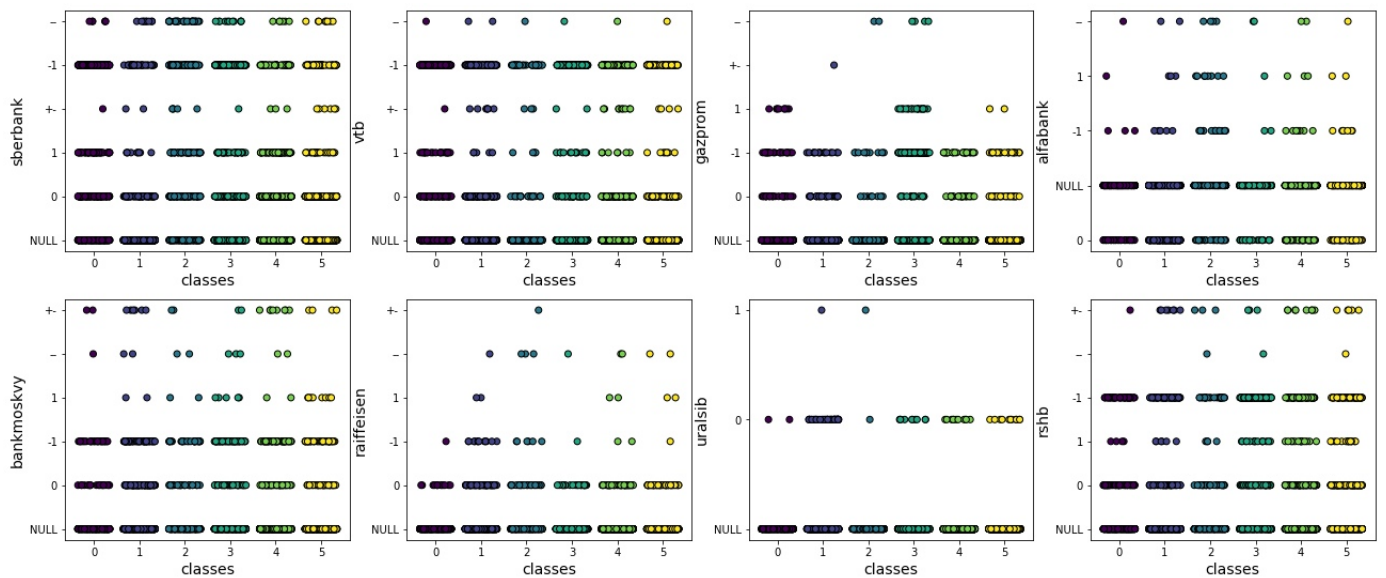


Figure 1 Bank comments clusters by topic model

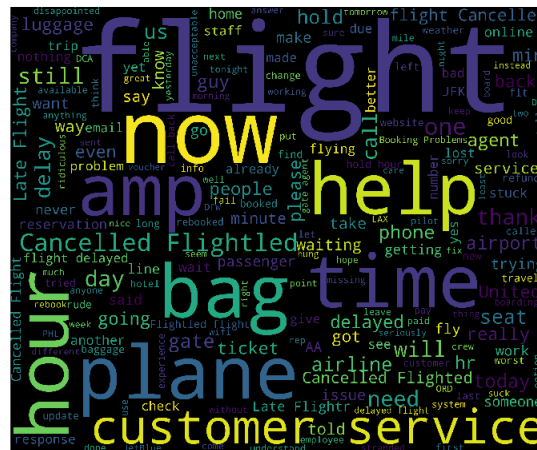


Figure 2 Negative wordcloud for aeroplanes



Figure 3 Positive wordcloud for aeroplanes