

# Neural-network Method for Determining Text Author's Sentiment to an Aspect Specified by the Named Entity

Aleksandr Naumov<sup>a,b</sup>, Roman Rybka<sup>a,b</sup>, Alexander Sboev<sup>a,b</sup>, Anton Selivanov<sup>a,b</sup> and Artem Gryaznov<sup>a,b</sup>

<sup>a</sup>National Research Centre "Kurchatov Institute", Moscow, Russia

<sup>b</sup>MEPhI National Research Nuclear University, Moscow, Russia

## Abstract

This study presents the approach to aspect-based sentiment analysis where a named entity of a certain category is considered as an aspect. Such task formulation is a novelty and opens up the opportunity to determine writers' attitudes to organizations and people considered in texts. This task required a dataset of Russian-language sentences where sentiment with respect to certain named entities would be labeled, which we collected using a crowdsourcing platform. Sentiment determination is based on a deep neural network with attention mechanism and ELMo language model for word vector representation. The proposed model is validated on available data on a similar task. The resulting performance (by the f1-micro metric) on the collected dataset is 0.72, which is the new state of the art for the Russian language.

## Keywords

text analysis, natural language processing, aspect based sentiment analysis, neural networks

## 1. Introduction

A relevant part of social monitoring is determining the sentiment of a text so as to identify its attitude to significant social events (aspects). Frequently, even one sentence contains several sentiment evaluations concerning various aspects of the text. For example, in the sentence: "Alex is an excellent worker, but the company Foo LLC, in which he works, poorly manages its staff", there are two named entities mentioned in different sentiment. The entity "Alex", of the category "Person", is used in positive sentiment, and the other entity "Foo LLC", of the category "Organization", in a negative one. This research proposes an approach to aspect-based sentiment determination in text with named entity pre-assigned as an aspect. Such task formulation is novel and opens up the opportunity to determine authors' attitudes to organizations and people considered in texts, which could be useful for social and political analysis.

There are several datasets in different languages available for aspect-based sentiment analysis task, including SemEval 2015 competition dataset [1], containing 830 reviews on three topics

---

*Russian Advances in Artificial Intelligence: selected contributions to the Russian Conference on Artificial intelligence (RCAI 2020), October 10-16, 2020, Moscow, Russia*

✉ Naumov-AV@nrcki.ru (A. Naumov); Rybka\_RB@nrcki.ru (R. Rybka); Sboev\_AG@nrcki.ru (A. Sboev); Selivanov\_AA@nrcki.ru (A. Selivanov); Gryaznov\_AV@nrcki.ru (A. Gryaznov)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

(laptops, restaurants, hotels) in English; SemEval 2016 [2], an extension of the previous one, containing more than 47000 sentences from reviews on seven topics (restaurants, laptops, mobile phones, telecommunications, digital cameras, hotels, museums) in eight languages, and in addition to those reviews, about 23000 more texts in six languages on three topics. The existing datasets in Russian language are the ones from the SentiRuEval competitions of years 2015 [3] and 2016 [4]. The first one contains 822 reviews of cars and restaurants and Twitter messages (23600 "tweets"). Overall, existing datasets are collected for specific topics and include aspect sets for some particular domains.

This research is based on a specially created dataset containing sentences with aspect-based sentiment labels. Crowdsourcing was used to extend the number of annotators, their markup was validated with a special procedure to provide annotation quality (see sec. 2). Analysis of the related works of the last few years [5, 6, 7] shows that algorithms based on complex topologies (including convolution, recurrent layers, attention mechanism) of deep learning models have a significant advantage over methods based on dictionaries, rules, and traditional machine learning methods. Therefore, this research uses neural network components in the solution development (see sec. 3). Results visualization uses Sankey diagrams, allowing to compare sentiment classes distribution over different named entities and various text sources (see sec. 5).

## 2. Dataset

### 2.1. Annotating Process

There are currently no datasets in Russian for setting up tools for solving problems of aspect analysis when the aspect is a named entity, so one has been gathered and annotated. For the formation of the corpus, we collected sentences in Russian from several sources: posts of the LiveJournal social network [8], texts of the online news agency Lenta.ru<sup>1</sup> and Twitter microblog posts [9]. A crowdsourcing platform<sup>2</sup> was used to annotate the sentences. Only Russian-speaking users of the 30% of the best performers among all active users of the platform by internal rating and over 18 years old were allowed into the annotation process. Before a platform user became an annotator, they underwent a training task, after which they were to mark 25 test samples, with more than 80% agreement with the annotation that we had performed ourselves. Upon successful completion of the training task, the user was allowed to complete the main tasks, 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 during the annotation process became less than 70% (including test and control samples), or if the percentage of correct answers to them was less than 66% over the last six control samples, then such annotator was blocked. A check was also performed on the number of consecutive identical labels and the time used for annotating the task. If the task was annotated too quickly (less than 30 seconds) or if there were many identical labels (more than eight), then such tasks were checked manually and removed from the sample if unfair labels

---

<sup>1</sup><https://github.com/yutkin/Lenta.Ru-News-Dataset>

<sup>2</sup><https://toloka.yandex.ru/>

were detected. Thus, each sentence was annotated at least three times.

## 2.2. Aspect-Based Sentiment Annotation

Generally, sentences do not always contain sentiment estimations and named entities. Thus, a preliminary selection of sentences for the subsequent annotation process is carried out. The selection criterion is the presence in the sentence of a named entity and one word from the sentiment words list. Named entities were extracted using a neural network model from the DeepPavlov library<sup>3</sup>, which is a State-of-the-art solution for the Russian language with an accuracy of 98.1% (f1-score metric \* 100%) obtained on the Collection3 dataset [10]. For filtering a sentence, we formed a list of sentiment words that based on dictionaries of opinionated words from domain-oriented Russian sentiment vocabularies of RuSentiLex [11]. About two thousand words were manually selected for positive sentiment, including: “joy”, “pleasure”, “cheerful”, etc.; and for a negative - about six thousand words, including: “enmity”, “ailment”, “grieve”, etc.

The annotators were asked to determine in what sentiment the author uses the named entity in the selected sentences (the classes of sentiment were “Positive”, “Neutral” and “Negative”). The sentence could not be marked with multiple tags. If the annotator was unable to unambiguously determine the sentiment class of the selected aspect, then that example was marked with the label “I find it difficult to determine” and, in the absence of other annotations, was not included in the resulting dataset. If a selected aspect was erroneously defined by a named entity, then such an example was marked as a “Wrong aspect” and was also not included in the final dataset. The final label for the sentence was selected on the basis of the aggregation of annotators labels by majority voting.

## 2.3. Summary of The Dataset

The aspect-based sentiment dataset contains 5552 unique sentences (1992 from Twitter, 2050 from the news site “Lenta.ru”, 1500 from the blog platform “Livejournal”). The resulting number of sentences for every presented sentiment label, as well as the number of unique named entities, are presented in Table 1.

**Table 1**

Summary of the dataset.

	Positive	Neutral	Negative	Number of unique named entities		
				Person	Organisations	Total
Twitter	977	510	510	1432	275	1818
Lenta	478	1653	472	1244	573	1817
LJ	834	905	366	1307	285	1592
Total	2289	3068	1348	3761	1068	4829

The number of unique entities was counted based on normal word forms without spaces and punctuation. The agreement was calculated as the average value of the ratio of the num-

---

<sup>3</sup><http://deeppavlov.ai/>

ber of answers for the selected sentiment label to the number of all answers for all entities. The agreement value was 0.84. The most similar datasets to the one collected from the point of aspect-based sentiment annotation for the Russian language are the datasets from SentiRuEval 2015-2016 competitions. However, the 2015 dataset contains 822 reviews (17000 particular entities) on two pre-defined topics (restaurants and cars) labeled with four sentiment classes (positive, negative, neutral, mixed). The 2016 dataset is more representative (approx. 23600 labeled entities), it contains labeling for pre-defined aspect list, which are possible not to be presented in sentence text. Therefore, the collected dataset is a significant extension of data available for sentiment analysis of Russian-language texts.

### 3. Method for Aspect-Based Sentiment Analysis

The proposed method is based on deep neural network with attention (IAN) [7], which solves a classification task. The architecture of the model consists of two parts: one processes the context for the target aspect, the other processes the words of the aspect itself. In our model, the context 1 is all the words of the sentence which contains a named entity, and the aspect 2 is the words that belong to the same named entity for which sentiment is determined.

$$context = [w_1, w_2, \dots, w_M], \quad (1)$$

$$aspect = [w_i, \dots, w_k], \quad (2)$$

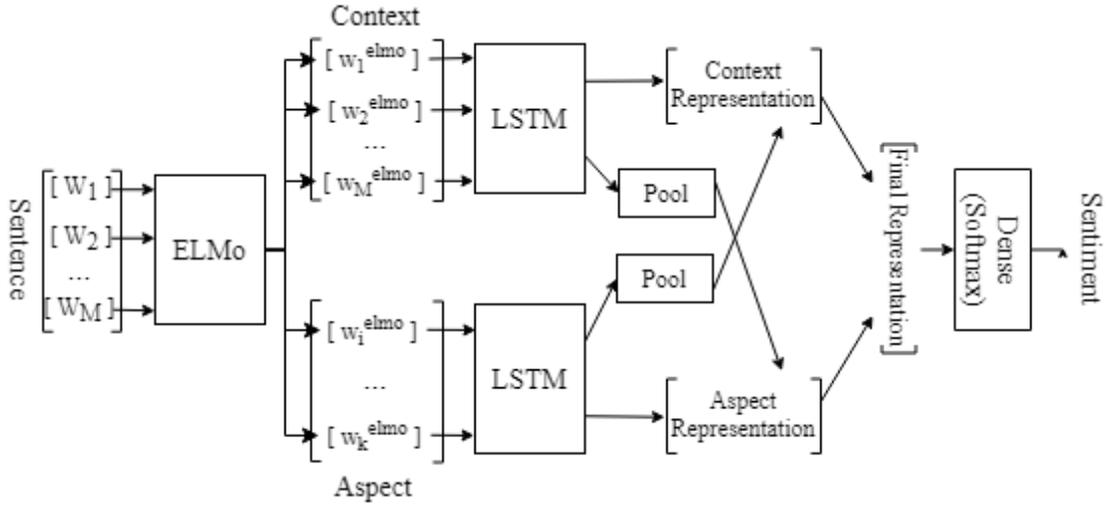
where  $M$  is the number of words in the sentence,  $i$  and  $k$  are the indices of the start and the end of the named entity, respectively. At the first step, the sentence words are vectorized using the bi-directional language model ELMo<sup>4</sup> [12], so that the representation of a word is the concatenation of representations from the hidden layers of the bidirectional language model.

Then, the vectors corresponding to the words of the aspect  $[w_i^{elmo}, \dots, w_k^{elmo}]$  and context  $[w_1^{elmo}, w_2^{elmo}, \dots, w_i^{elmo}, \dots, w_k^{elmo}, \dots, w_M^{elmo}]$  are selected. The resulting word embeddings of the aspect and context we feed into a recurrent neural network based on LSTM layers [13] to extract their internal states (“aspect representation” and “context representation”, respectively). After that, their average vectors are used to generate attention vectors. Next, the internal representations of the aspect and context are combined into the “Final Representation” vector, and the resulting vector is feed into a fully connected layer with the softmax activation function. Such implementation of the attention mechanism allows the target aspect and context to influence the formation of their internal representations in an interactive mode. The scheme of the proposed model architecture is presented in Fig. 1.

We evaluate the proposed model in terms of F1-macro and F1-micro scores (see section 4.1).

---

<sup>4</sup>[http://docs.deeppavlov.ai/en/master/features/pretrained\\_vectors.html#elmo](http://docs.deeppavlov.ai/en/master/features/pretrained_vectors.html#elmo)



**Figure 1:** Overview the architecture of the proposed model based on IAN.

## 4. Experiments

### 4.1. Metrics

To evaluate the performance of our models, we use the F1-measure metric as the evaluation score, as in the SentiRuEval 2015-2016 competitions.

$$Precision = TP / (TP + FP) \quad (3)$$

$$Recall = TP / (TP + FN) \quad (4)$$

$$F1 - measure = 2 * (Precision * Recall) / (Precision + Recall) \quad (5)$$

where TP is the number of true positives, FP is the number of false positives, FN is the number of false negatives.

In this case, the score is calculated in two variations (F1-macro and F1-micro). For macro-averaging the F1-measure calculation is averaged for each class separately, but for micro-averaging it is held for all examples together.

### 4.2. Experimental Results

In the experiments to determine the sentiment for a specified named entity, the deep neural network method proposed in this paper is validated on the original SentiRuEval-2015 competition dataset, and then trained and tested on the dataset collected in the current work.

The SentiRuEval-2015 competition dataset was originally split by the competition organizers into train and test set containing 5974 and 6615 samples respectively. Model training was based on examples of three classes: positive, negative, and mixed.

The best performance was achieved using the following hyperparameters for our model: Batch size – 4, Dropout – 0.3 (we add a dropout layer right before the recurrent LSTM layer), the number of neurons in the LSTM layers is 150, the learning rate is 0.01, the l2-regularization value is 0.001, the loss function is cross entropy [14].

Table 2 presents the model performance in terms of F1-micro and F1-macro scores. As a

**Table 2**

Model performance on the SentiRuEval-2015 dataset.

Model	Automobiles		Restaurants	
	F1-micro	F1-macro	F1-micro	F1-macro
SentiRuEval-2015 - baseline [3]	0.62	0.26	0.71	0.27
SentiRuEval-2015 - best [15]	0.74	0.57	0.82	0.55
<b>Our approach</b>	<b>0.79</b>	<b>0.6</b>	<b>0.85</b>	<b>0.58</b>

result, the method proposed in this work shows better accuracy compared to other solutions from the SentiRuEval-2015 competition [3], where the best result was shown by the gradient boosting model [15]. In that work, the authors was given a feature vector formed for each aspect using emotional lexicon compiled under some rather complex rules. These lexicons were formed for a particular dataset in each domain area separately using rules written manually by an expert.

The dataset collected in the current study was split into training and testing sets as 80% and 20% samples, respectively. The training set included 1081 named entities with negative sentiment, 1829 entities with a positive sentiment, and 2454 with neutral. The testing set included 460 positive entities, 267 negative entities, and 614 neutral entities. To evaluate the obtained model results, experiments were conducted with baseline methods that solve the usual problem of classifying sentences without paying any attention to the extracted named entity. These methods are built both on the rules using the vocabulary of emotive vocabulary, and with a classifier analyzing the entire sentence. To train such a classifier and select its hyperparameters, the AutoML method based on the TPOT library is used [16]. Thus, for comparison are used:

1. *Random*: Random definition of a label for each aspect;
2. *Lexicon*: This classifier is based on the positive and negative sentiment word lists that were in Section 2 used for pre-selecting sentences. The sentence is given a sentiment label that belongs to the dictionary the largest number of words from which are present in the sentence. If the number of words included in dictionaries of different sentiments is the same, then the label of the most representative sentiment of the corpus is put, that is, “positive” in this case. If the sentence doesn’t contain any words from the sentiment lists, then the sentiment of the sentence is considered neutral.
3. *TPOT(ELMo)*: This classifier is based on the TPOT software library. The average vectors of aspect words of the analyzed sentence obtained from the ELMo model are used as input features. The type of classifier and its parameters were selected automatically by the TPOT library.

The performance of the model on the proposed dataset, in comparison with the baseline methods, is presented in Table 3.

**Table 3**

Model performance on our dataset in terms of F1 scores.

Model	Twitter		LJ		Lenta. ru		All	
	micro	macro	micro	macro	micro	macro	micro	macro
Random	0.35	0.27	0.27	0.22	0.21	0.2	0.27	0.23
Lexicon	0.3	0.27	0.33	0.27	0.49	0.36	0.38	0.31
Tpot (ELMo)	0.63	<b>0.58</b>	0.57	0.56	0.74	0.7	0.65	0.56
<b>Our approach</b>	<b>0.64</b>	<b>0.58</b>	<b>0.67</b>	<b>0.66</b>	<b>0.78</b>	<b>0.72</b>	<b>0.72</b>	<b>0.7</b>

Besides, we have analyzed the ability of the model to classify aspects that are not present in the training set. The average performance across all sources has not changed. This confirms the effectiveness of the proposed approach for working with other named entities.

## 5. Results Visualization

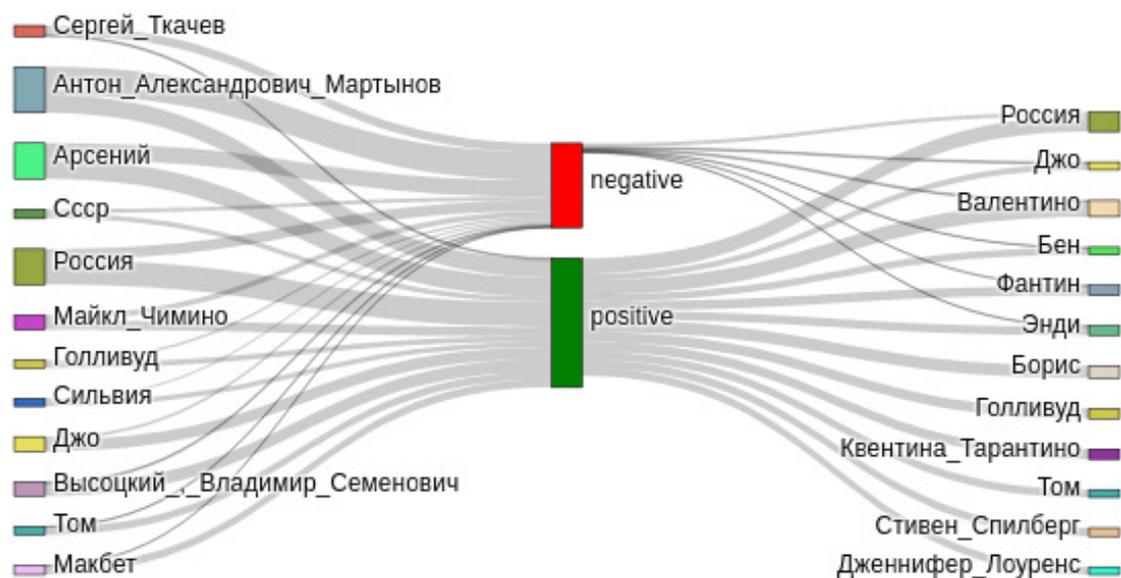
In this section, we present an example of a visualization of the results of aspect-based sentiment analysis. Experiments were conducted using Russian text corpus of the LiveJournal posts and the SCTM-ru dataset [17], compiled from articles from the Russian Wikinews website. For analysis, 47 news and 40 blog texts on the topic “cinema, oscar” was selected. These texts contain the following keywords: “film”, “role”, “cinema”, “oscar” etc.

Visualization of the results is carried out in the form of Sankey diagrams, which shows the frequency of named entities contained in different sentiment contexts (see Fig. 2). The figure 2 shows that authors of news articles have a more positive point of view in their publications, while LiveJournal posts authors often create negative context.

## 6. Future work

The main issues requiring further research are:

1. Reproducibility of results for other languages. This task is complicated by the fact that there are no labeled data sets where named entities are considered as aspects;
2. Verification and use of modern language models (for example BERT [18]), as well as other implementations of attention mechanisms;
3. Besides, it is planned to further develop the collected dataset, both from the side of increasing the number of examples and from the side of expanding sources and domain areas, which will make it possible to better assess the universality of the developed method;
4. Establishing identity between different spellings of the same entities to more accurately determine the integral assessment of their sentiment.



**Figure 2:** The frequency of mentions of named entities in negative and positive contexts for blog texts (left) and news (right).

## 7. Conclusion

The paper presents a deep-neural-network-based method for aspect-based sentiment analysis, where aspect is expressed by the named entity (organization or person), for textual data in Russian.

To solve the problem, a dataset of annotated sentences for several sources (blogs, microblogs, and news) were collected. The collected dataset is available for researchers upon request with <https://sagteam.ru/en> website. The developed method for building a dataset based on crowd-sourcing resources can be used to extend the dataset size and improve the performance of the proposed classifier. Also, the proposed method can be used in other domain areas to create labeled examples.

Evaluation of the model both on the open dataset from SentiRuEval-2015 competition and on the collected annotated corpus shows the efficiency of the developed solution. The resulting performance is a baseline for this type of task in Russian and allows one to provide aspect-based analysis with clear visualization of the results, an example of which is presented in the paper.

## Acknowledgments

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/>.

## References

- [1] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androustopoulos, Semeval-2015 task 12: Aspect based sentiment analysis, in: Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), 2015, pp. 486–495.
- [2] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androustopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, et al., Semeval-2016 task 5: Aspect based sentiment analysis, in: 10th International Workshop on Semantic Evaluation (SemEval 2016), 2016.
- [3] N. Loukachevitch, P. Blinov, E. Kotelnikov, Y. Rubtsova, V. Ivanov, E. Tutubalina, Sentirueval: testing object-oriented sentiment analysis systems in russian, in: Proceedings of International Conference Dialog, volume 2, 2015, pp. 3–13.
- [4] N. Lukashevich, Y. V. Rubtsova, Sentirueval-2016: overcoming time gap and data sparsity in tweet sentiment analysis, in: Komp’yuternaya lingvistika i intellektual’nyye tekhnologii, 2016, pp. 416–426.
- [5] B. Huang, K. M. Carley, Parameterized convolutional neural networks for aspect level sentiment classification, arXiv preprint arXiv:1909.06276 (2019).
- [6] P. Chen, Z. Sun, L. Bing, W. Yang, Recurrent attention network on memory for aspect sentiment analysis, in: Proceedings of the 2017 conference on empirical methods in natural language processing, 2017, pp. 452–461.
- [7] D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, arXiv preprint arXiv:1709.00893 (2017).
- [8] Rusprofiling lab 2017 rusprofiling corpus of russian texts, [online], ????, [Http://rusprofilinglab.ru/rusprofiling-atpan/corpus/](http://rusprofilinglab.ru/rusprofiling-atpan/corpus/).
- [9] Y. Rubtsova, Avtomaticheskoye postroyeniye i analiz korpusa korotkikh tekstov (postov mikroblogov) dlya zadachi razrabotki i trenirovki tonovogo klassifikatora. inzheneriya znaniy i tekhnologii semanticheskogo veba, Inzheneriya znaniy i tekhnologii semanticheskogo veba 1 (2012) 109–116.
- [10] V. Mozharova, N. Loukachevitch, Two-stage approach in russian named entity recognition, in: 2016 International FRUCT Conference on Intelligence, Social Media and Web (ISMW FRUCT), IEEE, 2016, pp. 1–6.
- [11] N. Loukachevitch, A. Levchik, Creating a general russian sentiment lexicon, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16), 2016, pp. 1171–1176.
- [12] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, arXiv preprint arXiv:1802.05365 (2018).
- [13] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.
- [14] R. Rubinstein, The cross-entropy method for combinatorial and continuous optimization, Methodology and computing in applied probability 1 (1999) 127–190.
- [15] J. Trofimovich, Comparison of neural network architectures for sentiment analysis of

russian tweets, in: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue, 2016, pp. 50–59.

- [16] T. T. Le, W. Fu, J. H. Moore, Scaling tree-based automated machine learning to biomedical big data with a feature set selector, *Bioinformatics* 36 (2020) 250–256.
- [17] S. Karpovich, The russian language text corpus for testing algorithms of topic model, *Intellektual'nyye tekhnologii na transporte* (2018).
- [18] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).