Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2023”

June 14–16, 2023

Simple Yet Effective Named Entity Oriented Sentiment Analysis

**Leonid Sanochkin[1,2], Angelina Bolshina[1], Kseniia Cheloshkina[1,2],
Daria Galimzianova[1,2], Aleksei Malafeev[1,2]**
[1]MTS AI,
[2]HSE,
{l.sanochkin, a.bolshina, k.cheloshkina, d.galimzianova, a.malafeev}@mts.ai

Abstract

Sentiment analysis, i.e. the automatic evaluation of the emotional tone of a text, is a common task in natural language processing. Entity-Oriented Sentiment Analysis (EOSA) predicts the sentiment of entities mentioned in a given text. In this paper, we focus on the EOSA task for the Russian news. We propose a text classification pipeline to solve this task and show its potential in such tasks. Moreover, in general, EOSA implies labeling both named entities and their sentiment, which can require a lot of annotator labour and time and, thus, presents a major obstacle to the development of a production-ready EOSA system. To help alleviate this, we analyse the potential of applying an Active learning approach to EOSA tasks. We demonstrate that by actively selecting instances for labeling in EOSA the annotation effort required for training machine learning models can be significantly reduced.

Keywords: Aspect-based sentiment analysis, Entity-oriented sentiment analysis, sentiment analysis, Active Learning

DOI: 10.28995/2075-7182-2023-22-459-468

Классификационный подход к анализу тональности именованных сущностей в новостных текстах

**Леонид Саночкин[1,2], Ангелина Большина[1], Ксения Челошкина[1,2],
Дарья Галимзянова[1,2], Алексей Малафеев[1,2]**
[1]МТС ИИ,
[2]НИУ ВШЭ,
{l.sanochkin, a.bolshina, k.cheloshkina, d.galimzianova, a.malafeev}@mts.ai

Аннотация

Автоматизированный анализ тональности текстов является одной из распространнных проблем автоматической обработки текстовой информации. В данной работе рассматривается оценка тональности по отношению к сущности в новостном тексте. Нами был предлоложен и протестирован подход, основанный на представлении данной задачи как задачи классификации. Кроме того, поскольку разметка данных для задач оценки тональности относительно сущности в тексте может быть трудоемким процессом, мы исследуем применимость активного обучения в данной задаче. Полученные результаты свидетельствуют о перспективности использования предложенного подхода в рамках активного обучения для задач оценки тональности относительно сущностей в тексте.

Ключевые слова: Анализ тональности текстов, тональность по отношению к сущности в тексте, активное обучение

1 Introduction

Nowadays, Aspect-based sentiment analysis (ABSA) is quite popular not only in the academic but also in the commercial sphere. Irrespective of the industry, it provides a fine-grained customer feedback analysis, offering valuable insights into the customer experience and helping to make data-driven decisions.

ABSA is a more fine-grained version of the classic sentiment analysis task that allows to obtain more detailed information from a text, which is more useful in real-life applications. The task of ABSA involves the extraction of various types of terms: 1) the aspect term (*a*); 2) the opinion term (*o*); 3) the aspect category (c) corresponding to the aspect term; 4) the sentiment polarity (s) for a given aspect term (Gao et al., 2022). ABSA can be divided into several sub-tasks based on the combinations of the identified terms. This article proposes an approach to solve the Entity-Oriented Sentiment Analysis (EOSA), which can be also referred to an Aspect-Category Sentiment Analysis.

Since it is necessary to label both entities and their sentiment inside the text, the costs of data annotation for entity-oriented sentiment analysis can hinder the practical application of such systems. Thus, we analyse the applicability of an Active learning (AL) pipeline for this problem, as described in the section 5.2. The results obtained show that our approach can be used for the active selection of instances to label and, thus, can be helpful in solving the EOSA task in a low-resource setting.

To summarize our contribution:

- We demonstrated that the entity-oriented sentiment analysis task can be efficiently solved with a naïve text classification pipeline;
- We addressed the problem of data shortage for such tasks and showed that by actively selecting examples to label, we can achieve comparable performance to the model trained on full data with a significantly smaller amount of labeled data;

2 Related work

Despite its high demand, ABSA task suffers from data scarcity, like many other NLP research areas. The survey (Chebolu et al., 2022) presents a comprehensive overview of available datasets for ABSA.

As mentioned above, ABSA consists of several sub-tasks, namely, aspect term and category identification, opinion term identification, and aspect sentiment classification. These tasks can be solved either separately or jointly. The former approach considers only one task at a time, e.g. (Li et al., 2020), (Xu et al., 2021a), (Ma et al., 2018). More often, studies focus on several subtasks simultaneously. All approaches differ in the number of the subtasks they solve. For example, the studies (He et al., 2019), (Dai et al., 2020), (Zhao et al., 2020) are devoted to the extraction of pairs of terms. Some papers identify triples in a text (Xu et al., 2020), (Wu et al., 2021). The approach described in (Cai et al., 2021) aims at quadruple extraction.

ABSA can be treated as classification, sequence tagging, machine reading comprehension tasks, or a generative problem. (Hu et al., 2019), (Jiang et al., 2019), and (Zhang and Qian, 2020) tackle ABSA as a classification problem. Some approaches transfer subtasks to the sequence tagging problem: (Li et al., 2019), (Chen and Qian, 2020), (Wu et al., 2021), (Xu et al., 2021b). (Yu et al., 2021), (Mao et al., 2021), (Liu et al., 2022), and (Chen et al., 2021) proposed to solve ABSA as a machine reading comprehension task. Generative frameworks are also used to solve ABSA subtasks: (Gao et al., 2022), (Zhang et al., 2021), (Yan et al., 2021), (Hosseini-Asl et al., 2022).

(Luo and Mu, 2022) studies EOSA in the news texts and proposes a Negative Sentiment Smoothing Model to address the multiple entity sentiment analysis problem. In (Fu et al., 2022), the problem of EOSA is studied on noisy data, obtained from automatic speech recognition tools.

3 Proposed approach

To address the problem of EOSA, we propose a text classification pipeline with an additional information on the analysed entity. We provide the model with additional information on the analysed entity by adding the exact entity string to the input token sequence with the separation token. Our approach is highly motivated by the success of solving question answering tasks with a machine reading comprehension pipeline, such as in (Devlin et al., 2018) and by the previously mentioned papers that reported solving ABSA with machine reading comprehension (Yu et al., 2021; Mao et al., 2021; Liu et al., 2022; Chen et al., 2021). In the section 5.2 we show with ablation studies that concatenating entity string with the input sequence is the key component that contributes greatly to the overall performance of the model for the EOSA task.

4 Dataset analysis

We evaluate our approach on the RuSentNE dataset (Golubev et al., 2023) created for the first competition in targeted sentiment analysis on named entities in Russian news. In the dataset, the named entities are already recognized and classified into the following labels: PERSON, ORGANIZATION, PROFESSION, COUNTRY, and NATIONALITY. The task is, for every sentence in the dataset, to assign a given entity one of the three sentiment classes: “positive” (“1”), “negative”(”-1”) or “neutral”(“0”). The sentences are not related, and there is always exactly one entity that needs to be labeled for sentiment. The dataset consists of three splits: training (6 637 examples, 15% negative / 72% neutral / 13% positive), validation (2 845 examples) and test (1947 examples). It is worth noting that, according to the survey (Chebolu et al., 2022), this dataset is one of the largest in terms of the number of entities.

As sentiment analysis is prone to be subjective, it is of interest here to investigate whether there are mislabeled examples or not. To get an understanding of how much data could be assigned wrong labels, we used the ”Dataset Cartography” method (Swayamdipta et al., 2020), which was shown to be effective in detecting labeling errors. This model-specific method assumes that every example in a dataset can be automatically categorized as belonging to one of the following groups: easy-to-learn examples (consistently labeled correctly by the model with high confidence), hard-to-learn examples (consistently mislabeled) and ambiguous examples (of high variability). We applied this method to the training set and built its data map. Results are presented in Figure 1. It can be clearly seen that this map has a low-density region of hard-to-learn examples, which means that the dataset has high annotation quality.

Nevertheless, since it was demonstrated that hard-to-learn examples tend to be labeling errors, it is worth taking a closer look at them. There are such 97 hard-to-learn examples out of 6 637 (1.5%) with a strong predominance of the positive class: the class balance is 27% / 24% / 49% in this subsample (”-1” / ”0” / ”1”), although in the full training sample the proportions are 15% / 72% / 13%. An inspection of hard-to-learn examples reveals some labeling errors is presented in the Table 4.

labeled as positive (but looks like at least neutral):
Подозреваемыми оказались два **студента**, каждому из которых по 21 году. (The suspects were two **students**, each of whom is 21 years old.)
Власти Парагвая объявили трёхдневный траур в связи с гибелью **политика**. (The Paraguayan authorities have declared three days of mourning in connection with the death of the **politician**.)
Кеплен вспоминает, что в ходе следствия было несколько нестыковок и пытается выяснить правду… (**Keplen** recalls that there were several inconsistencies during the investigation and is trying to find out the truth…)
labeled as negative:
Во время выступления **прокурора** он молча сидел, скрестив ноги и работая со своим планшетным компьютером. (During the **prosecutor’s** speech, he sat cross-legged in silence and worked with his tablet computer.)
Изучавший статую эксперт **Алессандро Мартелли** сказал: (The expert who studied the statue, **Alessandro Martelli**, said:)

Table 1: Examples of the label errors.

Thus, the dataset contains a small portion of mislabeled examples which were probably introduced by ambiguous annotation rules, as we further demonstrate in the model error analysis section.

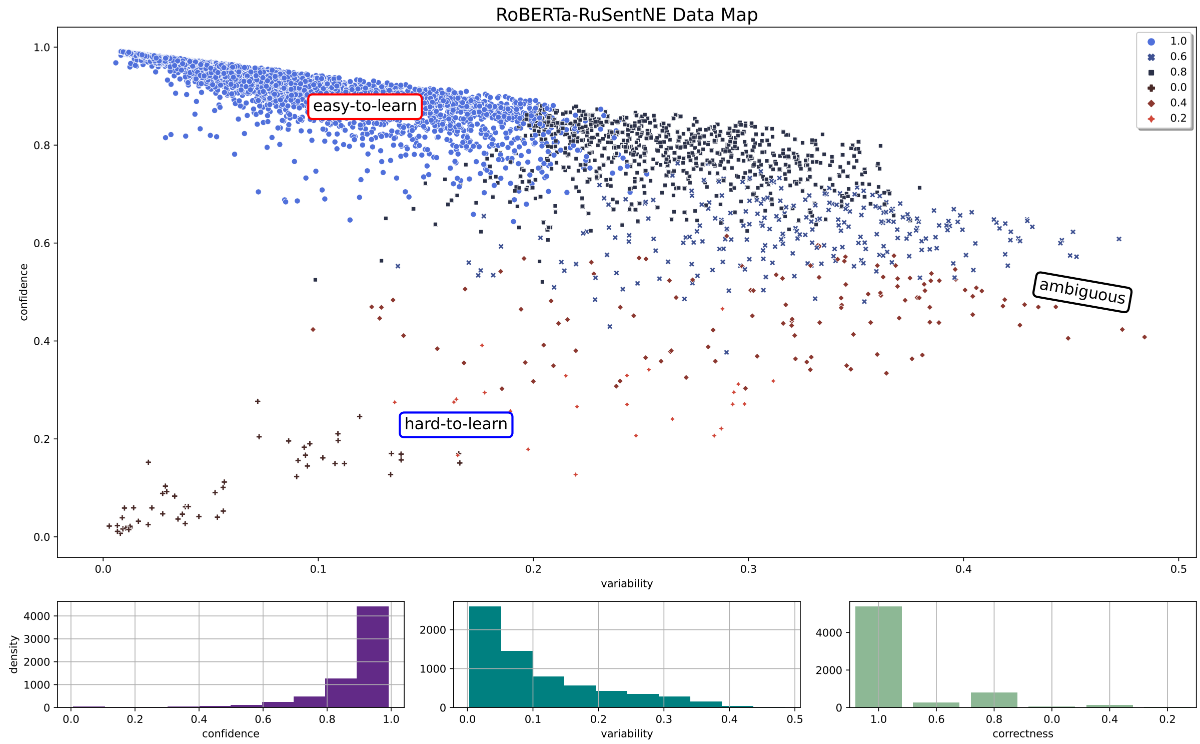

Figure 1: Dataset Cartography Map for RuSentNE.

5 Experiments

5.1 Experimental setup

The training data was randomly split into the training and validation parts in the 80/20 proportion. The provided results were computed on the test part of RuSentNE corpora via Codalab platform[1]. The competition uses a variant of a macro-$F1$ score ($F1_{pn}$), which is averaged over two sentiment classes: “positive” and “negative”. The class “neutral” is excluded because it is more relevant to extract opinions and sentiments. The results are averaged over five random seeds in order to report the standard deviation of the scores.

For active learning experiments, we used the classic simulated active learning experiment design (Settles and Craven, 2008; Shen et al., 2017). We emulated the AL annotation cycle starting with sampling from the dataset randomly and using this small portion of data as a seed for the construction of the initial acquisition model. On each iteration, a fraction of the top informative instances is sampled from the unlabeled pool by some query strategy. The selected instances are labeled according to the gold standard, then they are added to the training dataset and removed from the unlabeled pool for the following iterations. We used the following query strategies to score the informativeness of the unlabeled instances: Least Confidence (LC) (Lewis and Gale, 1994), Breaking Ties (BT) (Luo et al., 2004), Prediction entropy (PE) (Roy and Mccallum, 2001), and Contrastive Active Learning (CAL) (Margatina et al., 2021). We have not used some of the modern AL strategies, such as Batch Active Learning by Diverse Gradient Embeddings (BADGE) (Ash et al., 2020) and Batch Active learning via Information maTrices (BAIT) (Ash et al., 2021) due to their low computational performance and the fact that they cannot outperform the baseline strategies (such as LC) for a significant margin on a vast amount of datasets (Margatina et al., 2021; Tsvigun et al., 2022). For the successor model, we used the same model as for acquisition. To report standard deviations of the scores, we repeat the whole experiment five times

[1]https://codalab.lisn.upsaclay.fr/competitions/9538

with different random seeds. We sampled 2% of all training data (132 samples) and selected the same amount from the unlabeled pool on each iteration. We performed AL for 20 iterations.

As backbone models for our experiments, we used pretrained transformer models for Russian language: ruBert-base[2] and ruRoberta-large[3].

5.2 Results and discussion

Ablation studies In this section, we investigate different options for highlighting the specific entity of interest in the model input to perform entity-oriented sentiment analysis. We compared the following approaches:

1. Adding entity type info: concatenate the full sentence and the entity type string with the [SEP] separator. Input: "sentence [SEP] entityType".
2. Without entity information. Input: sentence.
3. In-text demonstration: add the [SEP] token before and after the entity text inside the sentence. Input: "sentenceStart [SEP] entity [SEP] sentenceEnd".
4. Our proposed approach: concatenate the full sentence and the entity string with the [SEP] separator. Input: "sentence [SEP] entity".

The results of the study are shown in the Table 5.2. The proposed approach outperforms the ones without proper information about an entity by a significant margin. However, pointing the entity inside the text leads to results within the confidence interval for the score.

Model	**Ours**	**Ablation 1**	**Ablation 2**	**Ablation 3**
ruBert-base	53.336±1.557	43.936±1.859	37.572±2.193	53.068±0.380
ruRoberta-Large	61.400±1.033	49.324±3.734	42.683±1.178	62.834±0.997

Table 2: Model performance.

We also include the performance of the baseline model and the top-performing approach from the competition in the Table 5.2. It can be seen that our approach, despite its simplicity, is quite competitive for the task of EOSA and has been outperformed by the top solution by a small margin.

Method	**F1**
Ours	62.92
Baseline	40.92
Best model	66.67

Table 3: Comparison with other methods.

Error analysis To perform error analysis, we used validation set labels obtained from five different seeds of our model, and compared them with the ground truth annotations. We also measured two types of agreement with Krippendorff's Alpha, which is a reliability coefficient ranging from -1 to 1 that can be used for two or more raters and categories, is applicable to many types of data and measurement scales, and has a number of other advantages (Krippendorff, 2011). First, we measured the agreement between all five seeds, which was very high: 0.79. This is expected, but we wanted to make sure that the model variations learn similar facts about the task from the training data regardless of the seed. Second, we also calculated the pairwise agreement between each seed and the ground truth. These ranged from 0.49 to 0.51: fairly close between the seeds and moderately high agreement with the ground truth.

Let us consider a few specific categories of errors. Out of 2845 examples, in 337 cases (about 11.5%) all five variations of our model yielded the same label, but different from the ground truth. In 46 of these, all seeds gave the opposite answer, i.e. either 1 instead of -1 or -1 instead of 1. More distributional

[2]https://huggingface.co/ai-forever/ruBert-base
[3]https://huggingface.co/ai-forever/ruRoberta-large

details are given in the Figure 2. Darker colors correspond to greater quantities of examples. GT stands for ”ground truth”. All percentages given are relative to the total number of examples in the validation set (2845).

<table>
<tr><td colspan="5">all examples</td></tr>
<tr><td colspan="5">2845 (100%)</td></tr>
<tr><td>all seeds agree with GT</td><td colspan="4">at least 1 seed disagrees with GT</td></tr>
<tr><td>2013 (71%)</td><td colspan="4">832 (29%)</td></tr>
<tr><td></td><td>some agree with GT</td><td colspan="3">all disagree with GT</td></tr>
<tr><td></td><td>448 (16%)</td><td colspan="3">384 (13%)</td></tr>
<tr><td></td><td></td><td>different answers</td><td colspan="2">same answer</td></tr>
<tr><td></td><td></td><td>47 (1.5%)</td><td colspan="2">337 (11.5%)</td></tr>
<tr><td></td><td></td><td></td><td>not opposite to GT</td><td>opposite to GT</td></tr>
<tr><td></td><td></td><td></td><td>291 (10%)</td><td>46 (1.5%)</td></tr>
</table>

Figure 2: Agreement of the models, trained on different random seeds.

It is noteworthy that when all five seeds disagree with the ground truth, in about 88% of the cases (337 vs 47) they are unanimous, i.e. yield the same answer. This might indicate labeling inconsistencies between the training and the test sets, at least in some cases. Consider the following examples:

- Пиночет совершил ошибку, приказав убить Неруду», — говорит Арайя. Pinochet made the mistake of ordering the death of Neruda,” says Araya.
  The ground truth label for the sentiment towards Neruda is questionably 1, while the five variations of the model unanimously suggest -1.

- Левая оппозиция желает проведения досрочных выборов, поскольку чувствует, что ветер успеха дует в ее паруса. The left opposition wants early elections because it feels that the wind of success is blowing in its sails. The ground truth label for “left opposition” is -1, while the model yields 1. Even if we accept that “positive” is a wrong answer, why is the ground truth answer not “neutral”?

These and other similar examples hint at the inherent difficulty and ambiguity of the targeted sentiment analysis task in the given setting. Indeed, the task description mentions that there are three possible sources of sentiment towards an entity: the author’s opinion, a quoted opinion of a third party, and an implicit opinion (Golubev et al., 2023). This raises some methodological concerns:

1. What if the author’s opinion and the quoted opinion are opposite, e.g. *They called my good friend Tom an idiot*. What is the sentiment towards Tom?
2. Is it possible to unambiguously define the implicit sentiment, when nothing but one sentence is given and we have no information about the author, the circumstances, etc.? For example, *Hitler came to power in 1933*. Should we consider the sentiment towards Hitler as negative because we know about his wrongdoings? But maybe the speaker is indeed pro-Hitler? Or is it a neutral context because the word choice is neutral? Or maybe ”coming to power” by itself can be considered as slightly positive?

This is further aggravated by the distribution of the ground truth labels in the test set: 2045 neutral examples (72%), 438 negatives (15%) and 362 positives (13%). There are fewer than 30% examples with non-neutral sentiment, and even some of these are questionable, as manual error analysis of the mislabeled examples shows. It is hard to quantify exactly how many of the sentences in the test set are mislabeled since there appears to be no obvious framework for unambiguous judgement on the 'correctness' of the labels, as discussed above.

On the Figure 3 is the confusion matrix for ground truth labels and model predictions aggregated by simple majority vote (there is always a majority since the number of seeds is greater than the number of possible labels and the number of seeds is odd).

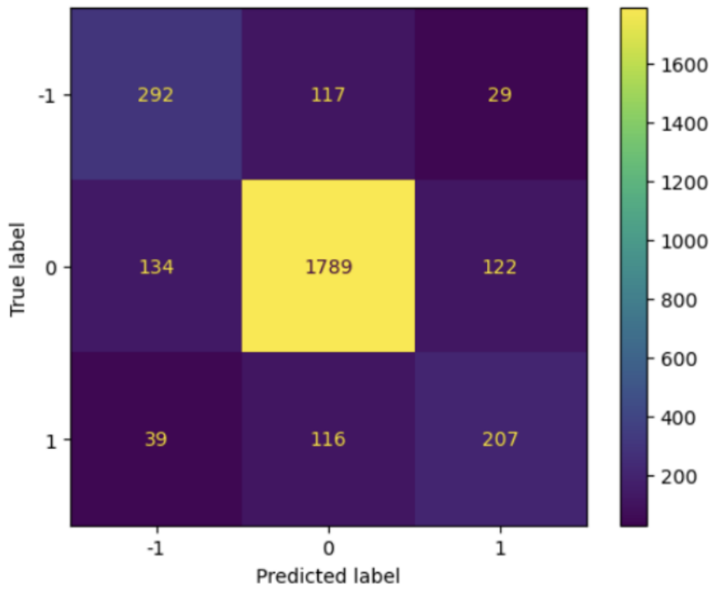

Figure 3: Confusion matrix.

As can be seen from the Figure 3, the model does not often confuse positive sentiment with negative (11% of all positive examples in the validation set) and negative for positive (6% of all negative examples in the validation set). However, there is a lot of confusion involving the neutral category (both type I and type II errors): 489 examples out of the total of 2845, or about 17%. This is understandable, as, firstly, the neutral category is the majority class, and secondly, it is easier to confuse neutral with positive / negative sentiment, rather than positive with negative or vice versa.

Active learning results The results of the best Active learning strategy are presented in Figure 4. It can be seen that the random sample selection baseline is outperformed by actively selecting samples according to an AL strategy. In our experiment, the best strategy for RuSentNE task was Breaking Ties, however, further research may be needed to determine the best query strategy and its hyperparameters for the EOSA tasks in general. Also, we plan to analyse the possibility of using smaller models as the acquisition model (without degrading successor performance) to make AL more efficient.

6 Conclusion

We analyzed the potential for solving EOSA tasks with a simple text classification pipeline and showed that our approach can be competitive in such tasks. Moreover, it can be easily adjusted to actively selecting instances for labeling. Our work demonstrates that active learning can be a promising approach for reducing the annotation effort in EOSA and improving the efficiency of the development of production-ready EOSA systems.

To further address the low-resource setting for EOSA tasks, we are looking forward to analysing the potential of applying few-shot methods for such tasks. Additionally, further research is needed on identifying the optimal hyperparameters of an AL pipeline.

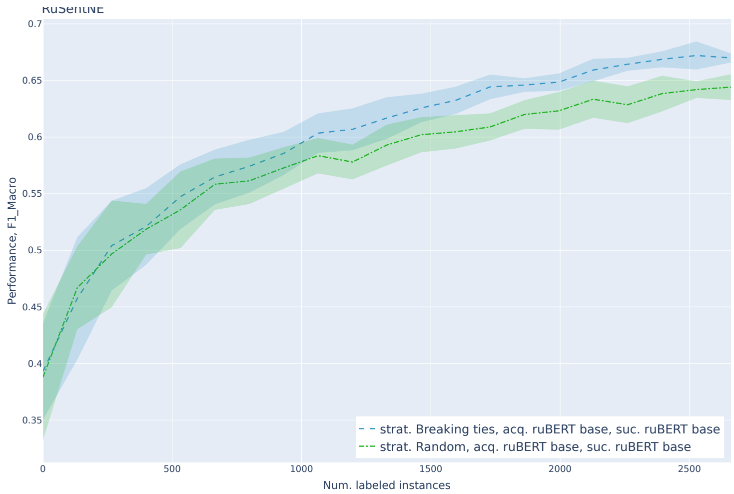

Figure 4: Active learning for RuSentNE.

References

Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. 2020. Deep batch active learning by diverse, uncertain gradient lower bounds. // *International Conference on Learning Representations*.

Jordan Ash, Surbhi Goel, Akshay Krishnamurthy, and Sham Kakade. 2021. Gone fishing: Neural active learning with fisher embeddings. // M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, *Advances in Neural Information Processing Systems*, volume 34, P 8927–8939. Curran Associates, Inc.

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. // *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, P 340–350, Online, August. Association for Computational Linguistics.

Siva Uday Sampreeth Chebolu, Franck Dernoncourt, Nedim Lipka, and Thamar Solorio. 2022. Survey of aspect-based sentiment analysis datasets. *arXiv preprint arXiv:2204.05232*.

Zhuang Chen and Tieyun Qian. 2020. Relation-aware collaborative learning for unified aspect-based sentiment analysis. // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, P 3685–3694, Online, July. Association for Computational Linguistics.

Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. // *Proceedings of the AAAI conference on artificial intelligence*, volume 35, P 12666–12674.

Zehui Dai, Cheng Peng, Huajie Chen, and Yadong Ding. 2020. A multi-task incremental learning framework with category name embedding for aspect-category sentiment analysis. // *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, P 6955–6965, Online, November. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar, Shayna Gardiner, Pooja Hiranandani, and Shashi Bhushan TN. 2022. Entity-level sentiment analysis in contact center telephone conversations. *arXiv preprint arXiv:2210.13401*.

Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao, and Weipeng Yan. 2022. LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. // *Proceedings of the 29th International Conference on Computational Linguistics*, P 7002–7012, Gyeongju, Republic of Korea, October. International Committee on Computational Linguistics.

Anton Golubev, Nicolay Rusnachenko, and Natalia Loukachevitch. 2023. RuSentNE-2023: Evaluating entity-oriented sentiment analysis on russian news texts. // *Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue"*.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, P 504–515, Florence, Italy, July. Association for Computational Linguistics.

Ehsan Hosseini-Asl, Wenhao Liu, and Caiming Xiong. 2022. A generative language model for few-shot aspect-based sentiment analysis. // *Findings of the Association for Computational Linguistics: NAACL 2022*, P 770–787, Seattle, United States, July. Association for Computational Linguistics.

Mengting Hu, Shiwan Zhao, Li Zhang, Keke Cai, Zhong Su, Renhong Cheng, and Xiaowei Shen. 2019. CAN: Constrained attention networks for multi-aspect sentiment analysis. // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, P 4601–4610, Hong Kong, China, November. Association for Computational Linguistics.

Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. A challenge dataset and effective models for aspect-based sentiment analysis. // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, P 6280–6285, Hong Kong, China, November. Association for Computational Linguistics.

Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability.

David D Lewis and William A Gale. 1994. A sequential algorithm for training text classifiers. // *SIGIR'94*, P 3–12. Springer.

Xin Li, Lidong Bing, Wenxuan Zhang, and Wai Lam. 2019. Exploiting BERT for end-to-end aspect-based sentiment analysis. // *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, P 34–41, Hong Kong, China, November. Association for Computational Linguistics.

Kun Li, Chengbo Chen, Xiaojun Quan, Qing Ling, and Yan Song. 2020. Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation. // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, P 7056–7066, Online, July. Association for Computational Linguistics.

Shu Liu, Kaiwen Li, and Zuhe Li. 2022. A robustly optimized BMRC for aspect sentiment triplet extraction. // *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, P 272–278, Seattle, United States, July. Association for Computational Linguistics.

Manman Luo and Xiangming Mu. 2022. Entity sentiment analysis in the news: A case study based on negative sentiment smoothing model (nssm). *International Journal of Information Management Data Insights*, 2(1):100060.

Tong Luo, K. Kramer, S. Samson, A. Remsen, D.B. Goldgof, L.O. Hall, and T. Hopkins. 2004. Active learning to recognize multiple types of plankton. // *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 3, P 478–481 Vol.3.

Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr.

Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. // *Proceedings of the AAAI conference on artificial intelligence*, volume 35, P 13543–13551.

Katerina Margatina, Giorgos Vernikos, Loïc Barrault, and Nikolaos Aletras. 2021. Active learning by acquiring contrastive examples. // *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, P 650–663, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Nicholas Roy and Andrew Mccallum. 2001. Toward optimal active learning through sampling estimation of error reduction. *Proceedings of the 18th International Conference on Machine Learning*, 08.

Burr Settles and Mark Craven. 2008. An analysis of active learning strategies for sequence labeling tasks. // *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, P 1070–1079, Honolulu, Hawaii, October. Association for Computational Linguistics.

Yanyao Shen, Hyokun Yun, Zachary Lipton, Yakov Kronrod, and Animashree Anandkumar. 2017. Deep active learning for named entity recognition. // *Proceedings of the 2nd Workshop on Representation Learning for NLP*, P 252–256, Vancouver, Canada, August. Association for Computational Linguistics.

Swabha Swayamdipta, Roy Schwartz, Nicholas Lourie, Yizhong Wang, Hannaneh Hajishirzi, Noah A. Smith, and Yejin Choi. 2020. Dataset cartography: Mapping and diagnosing datasets with training dynamics. // *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, P 9275–9293, Online, November. Association for Computational Linguistics.

Akim Tsvigun, Leonid Sanochkin, Daniil Larionov, Gleb Kuzmin, Artem Vazhentsev, Ivan Lazichny, Nikita Khromov, Danil Kireev, Aleksandr Rubashevskii, Olga Shahmatova, Dmitry V. Dylov, Igor Galitskiy, and Artem Shelmanov. 2022. ALToolbox: A set of tools for active learning annotation of natural language texts. // *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, P 406–434, Abu Dhabi, UAE, December. Association for Computational Linguistics.

Chao Wu, Qingyu Xiong, Hualing Yi, Yang Yu, Qiwu Zhu, Min Gao, and Jie Chen. 2021. Multiple-element joint detection for aspect-based sentiment analysis. *Knowledge-Based Systems*, 223:107073.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. // *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, P 2339–2349, Online, November. Association for Computational Linguistics.

Chi Xu, Hao Feng, Guoxin Yu, Min Yang, Xiting Wang, Yan Song, and Xiang Ao. 2021a. Discovering protagonist of sentiment with aspect reconstructed capsule network. // Christian S. Jensen, Ee-Peng Lim, De-Nian Yang, Wang-Chien Lee, Vincent S. Tseng, Vana Kalogeraki, Jen-Wei Huang, and Chih-Ya Shen, *Database Systems for Advanced Applications*, P 120–135, Cham. Springer International Publishing.

Lu Xu, Yew Ken Chia, and Lidong Bing. 2021b. Learning span-level interactions for aspect sentiment triplet extraction. // *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, P 4755–4766, Online, August. Association for Computational Linguistics.

Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. A unified generative framework for aspect-based sentiment analysis. // *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, P 2416–2429, Online, August. Association for Computational Linguistics.

Guoxin Yu, Jiwei Li, Ling Luo, Yuxian Meng, Xiang Ao, and Qing He. 2021. Self question-answering: Aspect-based sentiment analysis by role flipped machine reading comprehension. // *Findings of the Association for Computational Linguistics: EMNLP 2021*, P 1331–1342, Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Mi Zhang and Tieyun Qian. 2020. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. // *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, P 3540–3549, Online, November. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021. Towards generative aspect-based sentiment analysis. // *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, P 504–510, Online, August. Association for Computational Linguistics.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. SpanMlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, P 3239–3248, Online, July. Association for Computational Linguistics.