ABCD Team at ABSAPT 2024: Classification-Based versus Generation-Based Approach for Aspect-Based Sentiment Analysis in Portuguese

Dang Van Thin[1,2,*], Nguyen Tuan Kiet[1,2], Duong Ngoc Hao[1,2] and Ngan Luu-Thuy Nguyen[1,2]

[1]*University of Information Technology-VNUHCM, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam*

[2]*Vietnam National University, Ho Chi Minh City, Vietnam*

Abstract

This paper presents our solutions for Aspect Term Extraction (ATE) and Sentiment Orientation Extraction (SOE), the core tasks of the ABSAPT 2024 shared task. We investigate both classification-based and generation-based approaches using different pre-trained language models on the provided TripAdvisor review dataset for these two sub-tasks. Our system achieved the top ranking (1st place) for Task 1- Aspect Term Extraction and a top 3 ranking for Task 2 - Sentiment Orientation Extraction. Additionally, our work showcases the performance of these models for both tasks on Portuguese reviews, contributing valuable insights for further research in this area.

Keywords

Aspect-based Sentiment Analysis, Portuguese language, Classification-based approach, Generative-based approach,

1. Introduction

Aspect-based Sentiment Analysis (ABSA) is a fine-grained approach to sentiment analysis that addresses the limitations of traditional sentiment analysis, in which people's comments on specific aspects in reviews are overlooked.

Inspired by similar competitions such as SemEval [1, 2, 3] and EVALITA [4], the Aspect-Based Sentiment Analysis in Portuguese (ABSAPT) 2024 [5] at IberLEF 2024 [6] proposes to create an Aspect-Based Sentiment Analysis for TripAdvisor reviews written in Portuguese. This shared task involves two subtasks:

- **Aspect Term Extraction (ATE)**: give a set of reviews and identify all the aspects discussed within them.

IberLEF 2024, September 2024, Valladolid, Spain

*Corresponding author.

thindv@uit.edu.vn (D. V. Thin); 21521042@gm.uit.edu.vn (N. T. Kiet); haodn@uit.edu.vn (D. N. Hao); ngannlt@uit.edu.vn (N. L. Nguyen)

https://nlp.uit.edu.vn/ (D. V. Thin); https://nlp.uit.edu.vn/ (N. L. Nguyen)

0000-0001-8340-1405 (D. V. Thin); 0000-0002-3136-0661 (D. N. Hao); 0000-0003-3931-849X (N. L. Nguyen)

- **Sentiment Orientation Extraction (SOE)**: give a set of reviews and pre-identified aspects, and determine the sentiment (positive, neutral, negative) expressed towards each mentioned aspect.

The two subtasks of ABSA, Aspect Term Extraction (ATE) and Sentiment Orientation Extraction (SOE), are indeed interrelated and build upon each other. ATE acts as the foundation, identifying the specific aspects mentioned in the review. SOE then analyzes the sentiment expressed towards those identified aspects.

For instance, a review might mention ”the food was delicious” and ”the service was slow.” ATE would identify ”food” and ”service” as aspects. SOE would then analyze the surrounding text to determine the sentiment towards each aspect - positive for ”food” and negative for ”service.”

These subtasks are crucial for various applications. In the context of online reviews, they can help businesses pinpoint areas for improvement and focus customer satisfaction efforts on specific aspects that matter most to their customers. Furthermore, ABSA finds applications in social media analysis, product recommendation systems, and any domain where understanding user opinions on specific aspects is valuable.

By improving ABSA systems for Portuguese reviews, we can gain valuable insights from customer feedback on various platforms like TripAdvisor. This can benefit businesses by helping them understand customer satisfaction across different aspects of their services. Additionally, improved ABSA can be used for market research, product development, and enhancing customer service strategies.

This shared task encourages participants to develop and evaluate ABSA methods for Portuguese reviews, ultimately contributing to an enhanced understanding of how opinions are expressed in Portuguese.

2. Related work

Several recent approaches have addressed the challenges of Aspect Term Extraction (ATE) and Sentiment Orientation Extraction (SOE) using various techniques. Here, we explore some relevant works that participated in a similar shared task.

Team Deep Learning Brasil [7] leveraged a BERT-based model for the ATE task, while SOE was addressed as a sentence pair classification task. Another team, NILC [8], converted the data from document-level to sentence-level and utilized Conditional Random Fields to extract the aspect terms in the review. Instead of training supervised learning models, [9] adopted a simpler strategy relying on string matching techniques for Task 1. They also fine-tuned a BERT-base model for sequence classification to identify the sentiment for an aspect term. Similarly, the authors in [10] employed a BERT-base model fine-tuned for sequence classification to classify the sentiment polarity for aspect terms. Another work, [11], focused on enriching features through POS tagging, dependency parsing, and lemmatization to extract aspects. For SOE, they implemented a two-stage process: first extracting meaningful surrounding text for each aspect and then generating the sentiment polarity for each extracted aspect.

Beyond these works, several other approaches have explored different methodologies, including hybrid models and techniques for optimizing resource usage. [12] proposed a hybrid model combining rule-based and machine-learning methods for Aspect Term Extraction (ATE) and

Sentiment Orientation Extraction (SOE). They used predefined rules based on POS tagging to identify potential aspects, filtering out incorrect ones with a classifier, and employed Gradient Boosting with 800 Decision Trees for sentiment analysis. Meanwhile, [13] mixed transfer learning, zero-shot learning, and ONNX optimization to leverage the BERT-based Alberto [14] model efficiently. For ATE and SOE, they fine-tuned the model using Ktrain, and for sentiment analysis, they used zero-shot learning with AlBERTo embeddings and a BiLSTM classifier, achieving high inference speed with minimal CPU usage.

3. Methodology

3.1. Classification-based approach

For the classification-based approach, we utilize the fine-tuning pre-trained BERT-based language model method to address two sub-tasks. For the Aspect Term Extraction task, we treat it as the token-classification task and employ the BIO tagging scheme to represent the aspect terms in the review. Let X = $\{x_1, x_2, ..., x_n\}$ is the input review with n words. The purpose of this task is to predict a label sequence Y = $\{y_1, y_2, ..., y_n\}$ where each $y_i \epsilon \{B - Aspect, I - Aspect, O\}$ denotes the BIO tag for a token x_i. $B - Aspect$ indicates the beginning of a multi-word aspect term, $I - Aspect$ represents continuing words within the term, and O indicates the words outside any aspect term. We can formulate the Sentiment Orientation Extraction task as a sequence classification task. The model takes two inputs: the review and the aspect term. These are concatenated into a single sequence using special tokens, including [CLS] and [SEP]. The final representation becomes [CLS] review [SEP] aspect term [SEP]. The output is a one-hot vector with three dimensions corresponding to "negative", "neutral", and "positive" sentiment. Our purpose in this approach is to utilize the power of pre-trained BERT-based language models for the Portuguese language; therefore, we investigate different models as below:

- **BERTimbau** [15]: a pre-trained BERT model for Brazilian Portuguese that achieves state-of-the-art performances on three downstream NLP tasks: Named Entity Recognition, Sentence Textual Similarity and Recognizing Textual Entailment.
- **mDeBERTa_v3** [16]: is a multilingual version of DeBERTa architecture and was trained with CC100 multilingual data.
- **mBERT** [17]: a BERT model pre-trained on the top 104 languages with the largest Wikipedia using a masked language modelling (MLM) objective.
- **XLM-R** [18]: This is a multilingual model pre-trained on 2.5TB text corpora containing 100 languages, including Portuguese.
- **InfoXLM** [19]: an XLM-RoBERTa model focuses on maximizing the mutual information between text data in different languages and at various granularities.
- **XLM-ALign** [20]: an XLM-RoBERTa model specifically targets word-level alignment between parallel text corpora.

3.2. Generation-based approach

Instead of applying the classification-based approach, we also implement the generation-based approach to extract the aspect terms and sentiment polarity for a given aspect. Inspired by the previous works [21, 22, 23], we consider two sub-tasks as a conditional text generation task by utilizing the power of pre-trained generative language models. To do that, we transform the labels into a natural language string and fine-tune the encoder-decoder architecture for two sub-tasks. For Task 1 - Aspect Term Extraction, we use a special [sep] word to separate each aspect term mentioned in the review. For Task 2 - Sentiment Orientation Extraction, we convert the numeric sentiment classes (-1,0,1) as the corresponding sentiment polarity words (negativa, neutra, positiva) in Portuguese.

In order to train the models, we implement two different fine-tuning strategies, including the single-task and multi-task approach. For the single task, we train the models for each task separately. In contrast, we merge and train two sub-tasks simultaneously by adding the instruction prompt to distinguish two sub-tasks for the multi-task strategy. In this work, we employ the pre-trained language generative models mT5 [24] and mT0 [25] with two versions (base and large). This is a multilingual language model and supports the Portuguese language.

- **mT5** [24]: is pre-trained on a large multilingual dataset covering many languages. The model is trained using a unified text-to-text approach, where NLP tasks are framed as generating text from input text. This design choice simplifies using the model across different tasks and languages, providing a consistent methodology for various applications.
- **mT0** [25]: This is a multi-task prompted fine-tuning variant of mT5 (as cited in [24]) on various NLP tasks. The mT0 were trained on 30 new multilingual datasets, including one for sentiment analysis. Portuguese was one of the top 3 most-represented languages in the training set.

4. Experimental Setup

4.1. Data and Evaluation Metrics

Since the shared task only approves a final submission for both sub-tasks, we split the official training dataset into a new training set and a validation set with an 8:2 ratio for the development phase. For final submission, we will still train the best models on the official training set provided by the organizers. The official dataset is collected from the TripAdvisor reviews written in Portuguese, with 4,828 samples from 1,320 reviews for the train set; the test set comprises 283 unique review samples (task 1) and 1,176 samples from 282 reviews (task 2). As shown in Table 1, two sub-tasks have an imbalance problem. Our analysis revealed a class imbalance issue within the training dataset. Specifically, 69.28% of the aspect terms belong to the top 10 most frequent categories. This deviation is more imbalanced for the top 15 (78.66%) and top 20 (84.17%) aspects. These findings suggest potential challenges during model training. The model might prioritize learning frequently occurring aspects due to their over-representation, leading to a performance decline in identifying less frequent, yet potentially informative, aspects.

To evaluate our models' performance and address the class imbalance issue to some extent, we adopted different metrics for each sub-task. For Aspect Term Extraction (ATE), we employed the shared task metric of Accuracy along with Precision, Recall, and F1-score. These metrics provide a comprehensive view of the model's ability to identify both frequent and less frequent aspects in the reviews. In Sentiment Orientation Extraction (SOE), we utilized Balanced Accuracy

Table 1
Distribution of Top 20 most frequent aspect terms and sentiment in the training set.

ID	Aspect Term	Polarity			Total	ID	Aspect Term	Polarity			Total
		negative	neutral	positive				negative	neutral	positive	
1	hotel	116	135	651	902	11	limpeza	31	4	73	108
2	quarto	177	62	423	662	12	internet	28	9	54	91
3	localização	25	3	466	494	13	elevador	62	11	15	88
4	café da manhã	67	38	241	346	14	rua	17	32	36	85
5	atendimento	30	4	194	228	15	chuveiro	43	5	33	81
6	funcionários	18	2	140	160	16	cidade	1	50	21	72
7	preço	28	10	121	159	17	apartamento	7	14	42	63
8	recepção	30	12	91	133	18	lojas	1	14	33	48
9	serviço	27	17	87	131	19	aeroporto	0	44	4	48
10	cama	11	7	112	130	20	cassino	4	11	20	35

Table 2
Examples predicted by classification-based models

Predicted Aspects	True Aspects
['hotel', 'lojas', 'rua', 'quartos', 'elevador']	['rua']
['funcionários', 'quarto', 'café da manhã', 'padrão', 'localização']	['café da manhã', 'localização']

(the shared task metric) along with Accuracy, F1-micro, F1-macro, and F1-weighted. This combination assesses the model's performance across all sentiment classes, considering both overall accuracy and the balance between classes.

4.2. System Settings

In our classification approach, we experimented with all the BERT-based models mentioned above offered by HuggingFace, including both base and large versions. We limited the maximum input sequence length to 512 tokens. The learning rate was set at a low value of 3e-5, and we trained the models for 10 epochs. The batch size was chosen based on available resources and the specific model size. To optimize training time, we employed an Early Stopping Callback that automatically halts training when performance improvement plateaus. All experiments for this approach were conducted using NVIDIA P100 GPUs.

For the generation-based approach, we employ the mT5 language models with the base[1] and large[2] version downloaded directly from HuggingFace Hub. The maximum input and output length is set as 700 and 128 tokens. The learning rate is set to 3e-4, and the models will be trained for 20 epochs. We use the beam search as 5 to generate the target sequence. The batch size value is found automatically depending on the resources and size of the model. All the experiments were trained on the NVIDIA A100 with 80GB GPU.

Table 3
Results on the development set for Task 1: Aspect Term Extraction

Approach	Model	Version	Evaluation Metrics			
			Accuracy	Precision	Recall	F1-score
Classification-based	BERTimbau	base	92.57	77.72	92.57	84.50
		large	92.57	79.84	92.57	85.74
	mBERT	base	94.21	75.56	94.21	83.86
	XLM-Align	base	93.77	75.88	93.77	83.88
	Info-XLM	base	95.30	76.04	95.30	84.59
		large	93.66	78.78	93.66	85.58
	mDeBERTa_v3	base	93.32	77.54	93.32	84.70
	XLM-R	base	94.75	76.14	94.75	84.43
		large	96.17	76.40	96.17	85.16
Generation-based	single mT5	base	80.45	79.85	80.45	80.15
		large	81.95	80.40	81.95	81.17
	single mT0	base	80.66	78.31	80.66	79.47
		large	84.31	79.29	84.31	81.73
	multi-task mT5	base	80.77	78.99	80.77	79.87
		large	80.67	79.89	80.67	80.28
	multi-task mT0	base	80.88	80.53	80.88	80.71
		large	82.06	82.25	82.06	82.15

5. Main results

In Table 3 and Table 4, we present the performance of two approaches with different models on the development set for Task 1 and Task 2, respectively.

For ATE, classification-based approaches consistently outperformed generation-based approaches. This suggests that directly classifying aspects within the review text might be more effective for this task compared to attempting to generate them from scratch. It's possible that the complexity of aspect term variations and their dependencies on context are not fully captured by the generation process in this setting. Interestingly, the gap in performance between classification and generation-based approaches was narrower for metrics like precision and F1-score compared to accuracy. This might be because classification models tend to detect overly inclusive aspects, including all potential aspects the model identifies during evaluation, as shown in Table 2. While this can inflate accuracy, it might not reflect a true understanding of the most relevant aspects of the review.

For SOE, the results revealed that generation-based approaches achieved competitive performance, even surpassing most classification-based models in several metrics. This suggests that these approaches can effectively capture sentiment orientation in Portuguese reviews.

While classification-based models achieved high overall accuracy, they might exhibit limitations in capturing the nuances of sentiment compared to generation-based approaches. This is supported by potential biases towards certain sentiment classes observed in some classification models.

[1] https://huggingface.co/google/mt5-base
[2] https://huggingface.co/google/mt5-large

Table 4
Results on the development set for Task 2: Sentiment Orientation Extraction

Approach	Model	Version	Evaluation Metrics				
			Accuracy	Micro F1	Macro F1	Weighted F1	Balanced Acc
Classification-based	BERTimbau	base	69.56	69.56	52.78	67.06	53.48
	mBERT	base	73.18	73.18	50.21	67.60	54.81
	XLM-Align	base	73.91	73.91	57.65	71.09	56.71
	Info-XLM	base	73.80	73.80	49.98	67.64	54.69
	mDeBERTa_v3	base	81.57	81.57	76.27	82.14	78.24
	XLM-R	base	71.84	71.84	54.14	68.97	55.27
Generation-based	single mT5	base	70.08	70.08	47.53	65.07	50.39
		large	69.15	69.15	53.41	67.30	53.48
	single mT0	base	81.25	81.25	73.64	80.73	73.42
		large	82.40	82.40	74.99	81.90	74.21
	multi-task mT5	base	80.23	80.23	72.60	79.90	71.87
		large	81.37	81.37	73.57	80.85	72.65
	multi-task mT0	base	81.99	81.99	74.42	81.50	73.88
		large	82.92	82.92	75.67	82.60	75.00

Table 5
The official ranking of our system for two sub-tasks.

Team	Task 1				Task 2				
	Precision	Recall	F1-score	Top	Balanced Acc	Precision	Recall	F1-score	Top
Emerson	6.57	6.68	6.68	2	78.40	76.80	78.40	77.47	1
TeamUFPR	-	-	-	-	65.19	65.34	65.19	65.34	2
Ours	85.52	73.04	63.73	1	57.13	56.83	56.93	56.83	3

To further analyze the performance of our models, we also draw confusion matrices of the best models from each approach for Task 2's results presented in Figure 1. These confusion matrices provide insightful comparisons of their performance across negative, neutral, and positive sentiment categories. The generation-based approach demonstrates a high precision in classifying positive sentiments (0.92), while its performance for neutral (0.54) and negative (0.79) sentiments is comparatively moderate. Misclassification is most evident in neutral sentiments being predicted as positive (0.32). In contrast, the classification-based approach exhibits a more balanced performance with a precision of 0.86 for negative sentiments, 0.64 for neutral, and 0.85 for positive sentiments. Despite this balance, there is still a significant misclassification rate between neutral and positive sentiments. These results indicate that while the generation-based model excels in identifying positive sentiments, the classification-based model offers more consistent performance across all categories, highlighting the strengths and weaknesses of each approach in sentiment orientation extraction.

Table 5 presents the official ranking results of Task 1 - ATE and Task 2 - SOE. Our team secured first place in Task 1, achieving outstanding metrics with a precision of 85.52, recall of 73.04, and an F1-score of 63.73, surpassing all other competitors. For Task 2, our results in Task 2 include a balanced accuracy of 57.13, precision of 56.83%, recall of 56.93%, and an F1-score of 56.83%. While our system achieved impressive results in the shared task, a more nuanced analysis is necessary, given the limited number of participating teams. In Task 1, our system outperformed the only other competitor, demonstrating superior performance. However, the presence of only two teams makes it challenging to assert the overall efficacy and robustness of

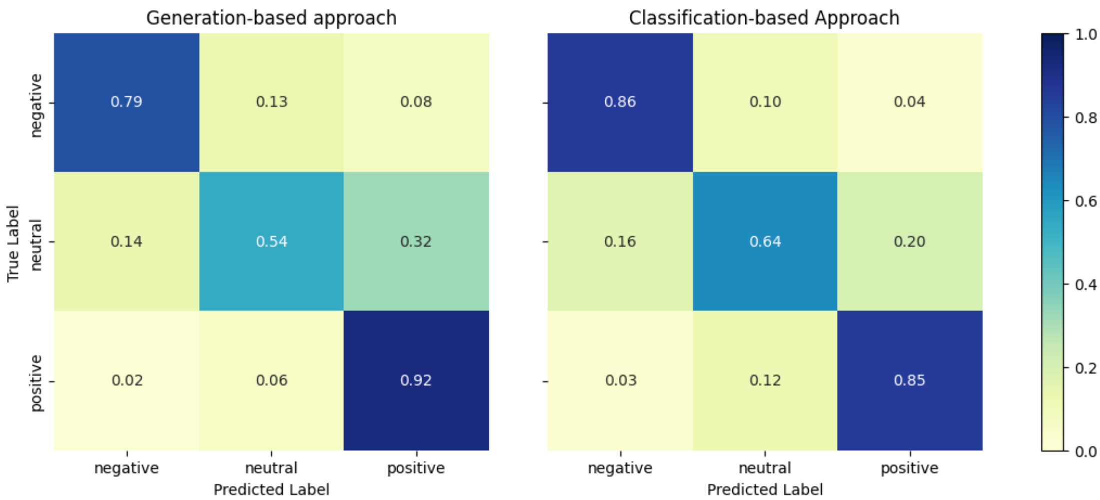

Figure 1: Confusion Matrices of the Best Models from each approach on Task 2: Sentiment Orientation Extraction

our approach, as the competition was not extensive. Similarly, in Task 2, the small number of participants restricts the breadth of comparative analysis. Therefore, while our results indicate a strong performance relative to the available competitors, further validation against a larger and more diverse set of systems would be necessary to demonstrate the superiority of our approach conclusively.

6. Conclusion and Future Work

In this work, we explored two approaches for aspect term extraction and sentiment orientation extraction in the ABSAPT 2024 shared task. We fine-tuned various BERT-based architectures (BERT, XLM-RoBERTa, RoBERTa, DeBERTa) for a classification-based approach. Additionally, we investigated the effectiveness of generative models (mT5, mT0) with single-task and multi-task strategies for a generation-based approach. Our team achieved notable results in the competition, securing first place in Task 1 and reaching the top 3 in Task 2. These rankings underscore the strength and versatility of our approaches, although further validation is needed against a larger and more diverse set of systems. Future efforts can investigate ensemble methods combining classification and generation approaches and develop techniques for solving class imbalance issues. Additionally, future work should address capturing implicit aspects and developing explainable models to build trust and enable error analysis.

Acknowledgments

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number C2024-26-02. Dang Van Thin was funded by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2023.TS117.

References

[1] I. Pavlopoulos, Aspect based sentiment analysis, Athens University of Economics and Business (2014).

[2] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, SemEval-2015 task 12: Aspect based sentiment analysis, in: P. Nakov, T. Zesch, D. Cer, D. Jurgens (Eds.), Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, 2015, pp. 486–495.

[3] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, 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: ProWorkshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, 2016, pp. 19–30.

[4] L. De Mattei, G. De Martino, A. Iovine, A. Miaschi, M. Polignano, G. Rambelli, et al., Ate absita@ evalita2020: Overview of the aspect term extraction and aspect-based sentiment analysis task, in: CEUR WORKSHOP PROCEEDINGS, volume 2765, CEUR-WS, 2020, pp. 67–74.

[5] A. G. Gabriel, T. B. Alexandre, P. L. Emerson, A. d. F. Larissa, B. C. Ulisses, Overview of absapt at iberlef 2024: Overview of the task on aspect-based sentiment analysis in portuguese, Procesamiento del Lenguaje Natural 73 (2024).

[6] L. Chiruzzo, S. M. Jiménez-Zafra, F. Rangel, Overview of IberLEF 2024: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS.org, 2024.

[7] J. R. S. Gomes, E. A. S. Garcia, A. F. B. Junior, R. C. Rodrigues, D. F. C. Silva, D. F. Maia, N. F. F. da Silva, A. d. S. Soares, et al., Deep learning brasil at absapt 2022: Portuguese transformer ensemble approaches, arXiv preprint arXiv:2311.05051 (2023).

[8] M. T. Machado, T. A. S. Pardo, Nilc at absapt 2022: Aspect extraction for portuguese, in: Proceedings, 2022.

[9] F. A. R. Neto, R. F. de Sousa, R. L. de Sales Santos, R. T. Anchiêta, R. S. Moura, Team piln at absapt 2022: Lexical and bert strategies for aspect-based sentiment analysis in portuguese. (2022).

[10] T. Heinrich, F. Marchi, Teamufpr at absapt 2022: Aspect extraction with crf and bert. (2022).

[11] F. M. Assi, G. B. Candido, L. N. dos Santos Silva, D. F. Silva, H. de Medeiros Caseli, Ufscar's team at absapt 2022: Using syntax, semantics and context for solving the tasks. (2022).

[12] A. S. F. Mele, G. Vettigli, Sentna@ ate absita: Sentiment analysis of customer reviews using boosted trees with lexical and lexicon-based features, Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020), Online. CEUR. org (2020).

[13] M. Bennici, ghostwriter19@ ate_absita: Zero-shot and onnx to speed up bert on sentiment analysis tasks at evalita 2020, EVALITA Evaluation of NLP and Speech Tools for Italian-December 17th, 2020 (2020) 80.

[14] M. Polignano, P. Basile, M. De Gemmis, G. Semeraro, V. Basile, et al., Alberto: Italian

bert language understanding model for nlp challenging tasks based on tweets, in: CEUR workshop proceedings, volume 2481, CEUR, 2019, pp. 1–6.
[15] F. Souza, R. Nogueira, R. Lotufo, BERTimbau: pretrained BERT models for Brazilian Portuguese, in: 9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear), 2020.
[16] P. He, J. Gao, W. Chen, Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, 2021. arXiv:2111.09543.
[17] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186.
[18] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 8440–8451.
[19] Z. Chi, L. Dong, F. Wei, N. Yang, S. Singhal, W. Wang, X. Song, X.-L. Mao, H. Huang, M. Zhou, InfoXLM: An information-theoretic framework for cross-lingual language model pre-training, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 3576–3588.
[20] Z. Chi, L. Dong, B. Zheng, S. Huang, X.-L. Mao, H. Huang, F. Wei, Improving pretrained cross-lingual language models via self-labeled word alignment, in: 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), Association for Computational Linguistics, Online, 2021, pp. 3418–3430.
[21] S. U. S. Chebolu, F. Dernoncourt, N. Lipka, T. Solorio, Exploring conditional text generation for aspect-based sentiment analysis, in: K. Hu, J.-B. Kim, C. Zong, E. Chersoni (Eds.), Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation, Association for Computational Lingustics, Shanghai, China, 2021, pp. 119–129.
[22] H. Yan, J. Dai, T. Ji, X. Qiu, Z. Zhang, A unified generative framework for aspect-based sentiment analysis, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), 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), Association for Computational Linguistics, Online, 2021, pp. 2416–2429.
[23] D. Van Thin, N. L. T. Nguyen, Aspect-category based sentiment analysis with unified sequence-to-sequence transfer transformers, VNU Journal of Science: Computer Science and Communication Engineering 39 (2023).
[24] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mT5: A massively multilingual pre-trained text-to-text transformer, in: K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, Y. Zhou (Eds.), Proceedings of the 2021 Conference of the North American

Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 483–498.
[25] N. Muennighoff, T. Wang, L. Sutawika, A. Roberts, S. Biderman, T. Le Scao, M. S. Bari, S. Shen, Z. X. Yong, H. Schoelkopf, X. Tang, D. Radev, A. F. Aji, K. Almubarak, S. Albanie, Z. Alyafeai, A. Webson, E. Raff, C. Raffel, Crosslingual generalization through multitask finetuning, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 15991–16111.