Aspect-Based Sentiment Analysis of Clothing Reviews in Vietnamese E-commerce*

Pham Quoc-Hung[1], Dinh Van-Dan[1], Le Huu-Loi[1], Le Thi-Viet-Huong[3], Nguyen Thu-Ha[1], Phan Xuan-Hieu[2], Nguyen Minh-Tien[1], Pham Ngoc-Hung[4],

[1]Hung Yen University of Technology and Education, Hungyen, Vietnam
[2]VNU University of Engineering and Technology, Hanoi, Vietnam
[3]National Hospital of Obstetrics and Gynecology, Hanoi, Vietnam
[4]Phenikaa University, Hanoi, Vietnam
quochungvnu@gmail.com

Abstract

Significant advancements have been achieved in sentiment analysis; however, aspect-based sentiment analysis (ABSA) remains underexplored in the Vietnamese language despite its vast potential across various natural language processing applications, including 1) monitoring sentiment related to products, movies, and other entities; and 2) enhancing customer relationship management models. A huge number of reviews are generated on e-commerce platforms, and analyzing them in depth brings a lot of helpful information to users. This paper presents the first standard Vietnamese dataset for the clothing reviews domain. Specifically, we create a new Vietnamese dataset, ViCloABSA, as a new benchmark based on a strict annotation scheme for evaluating aspect-based sentiment analysis. The proposed dataset comprises 7,000 human-annotated comments with five aspect categories and three polarity labels for clothes collected from e-commerce platforms. The dataset is freely available for research purposes [1]. We experiment with this dataset using strong baselines and report error analysis. The evaluation results show that a model based on large language models is superior to other existing works.

1 Introduction

Aspect-based sentiment analysis is challenging in natural language processing (NLP) due to its need for fine-grained sentiment classification, accurate aspect extraction, and contextual understanding. The complexity of the task is heightened by factors such as sparse data, interdependencies among aspects, and the dynamic nature of language (Liu, 2020). With the boom of e-commerce, customers generate a large number of user feedback reviews on these platforms every day. These reviews are effective for customers, manufacturers, and service providers.

People are very interested in costumes, so e-commerce platforms sell many of these products. When buying a set of clothes, customers often find out information about some aspects of the product, such as material, design, price, and more. Thanks to this, it is possible to conduct some analysis to understand customers' attitudes towards clothes deeply. This rationale underpins our decision to select clothing reviews for constructing a dataset that addresses the Aspect-Based Sentiment Analysis challenge within the context of e-commerce reviews.

While the ABSA task has shown encouraging results in English across various numerical datasets, much research hasn't been done on it in Vietnamese, especially for clothing products. This paper fills the gap by investigating the capability of five advanced methods for ABSA in Vietnamese with a new dataset for clothing. In summary, this paper presents two main contributions.

- It introduces a Vietnamese dataset focusing on clothing reviews from e-commerce platforms, specifically designed for the ABSA tasks.
- It conducts a comprehensive evaluation of robust baseline models tailored to ABSA tasks.

2 Related Work

ABSA datasets have significantly contributed to the progression of sentiment analysis research, particularly in the context of product reviews. Various datasets have driven recent advancements in ABSA. Notable datasets include the SemEval-2014 Restaurant and Laptop datasets (Pontiki et al., 2014), which were early benchmarks for ABSA tasks, covering restaurant and laptop product reviews. The SentiHood dataset (Saeidi et al., 2016) extended ABSA to location-based sentiment analy-

*The corresponding author is Pham Ngoc-Hung.
[1]https://github.com/quochungvnu24/ViCloABSA

sis in a real-world context. The MultiAspect Multi-Sentiment (MAMS) dataset was presented by Jiang et al. (2019), in which each sentence contains multiple aspects with different sentiment polarities. Expanding ABSA to non-English contexts, the Chinese Review Datasets (ASAP) by Bu et al. (2021) provided a crucial resource for Chinese product reviews. Most recently, Xu et al. (2023) presented a Diversified Multi-domain Dataset For Aspect Sentiment Triplet Extraction (DMASTE), manually annotated to better fit real-world scenarios by providing more diverse and realistic reviews.

In Vietnamese, several datasets for sentiment analysis across various domains are available. For instance, Tran and colleagues from VNUHCM - University of Information Technology (Tran et al., 2022) introduced a Vietnamese dataset specifically tailored for assessing lipstick products within the context of ABSA. Luc Phan et al. (2021) presented the UIT-ViSFD dataset, a Vietnamese Smartphone Feedback Dataset comprising 11,122 human-annotated comments related to mobile e-commerce. Furthermore, Nguyen et al. (2018) made public the SA-VLSP2018 dataset, designed for ABSA tasks focusing on the restaurant and hotel domains. Additionally, Nguyen and collaborators (Van Nguyen et al., 2018) released the UITVSFC dataset, which is centered on student feedback analysis. These datasets have facilitated extensive research and model development in ABSA tasks. However, to the best of our knowledge, there has been no Vietnamese dataset on clothing reviews for the ABSA task yet. It motivates the creation of a new dataset, ViCloABSA, for this problem.

3 Dataset

We have built a comprehensive Vietnamese dataset comprising customer reviews related to clothing products tailored specifically for the ABSA task. This dataset contains a collection of 7,000 reviews acquired from Shopee[2] and Lazada[3], which are two popular e-commerce platforms in Vietnam.

It encompasses two sub-tasks: aspect detection and sentiment classification. In the aspect detection sub-task, our focus is directed toward identifying and categorizing aspects discussed within the feedback reviews. These aspects encompass five categories: MATERIAL, DESIGN, PRICE, SERVICE, and GENERAL, each meticulously defined as presented in Table 1. These aspects are selected based on their popularity and importance in clothing reviews, facilitating a more comprehensive analysis and providing detailed, helpful information for businesses and customers. Additionally, the dataset also entails the second sub-task of classifying the sentiment polarity of these aspects as either *positive*, *negative*, or *neutral*.

3.1 Data Collection Process

The data collection process was systematically executed through the acquisition of Vietnamese product reviews pertaining to T-shirts from two prominent e-commerce platforms, Shopee and Lazada, which enable customers to write fine-grained reviews regarding the T-shirts they have purchased or utilized.

To collect review data, we utilized a combination of web scraping tools and APIs. Our data collection process is conducted on the basis of respecting customer privacy and complying with data ownership regulations. Specifically, we collected product reviews and ensured that all personal information remained anonymous. In the reviews, users give positive, neutral, or negative opinions on many aspects, such as MATERIAL, DESIGN, PRICE, SERVICE, and GENERAL.

3.2 Data Annotation Process

Following the completion of data collection, the subsequent phase involved the meticulous annotation of the acquired dataset utilizing the *Label-Studio* tool. Two stages made up the data annotation process: Phase 1 concentrated on combining guidelines, while Phase 2 observed annotators utilizing the established guideline to annotate the remaining samples, ensuring a systematic and consistent approach throughout the entire annotation process.

In the initial phase, a stratified random sampling method selected 200 reviews, which were divided into two segments for systematic annotation. The goal was to identify aspects within the reviews and assess the associated sentiment. In the annotation phase, two annotators participated, and their labeling outputs were compared using Cohen's Kappa coefficient to measure agreement scores. This process was repeated to optimize the Kappa score and create a comprehensive annotation guideline. After achieving high inter-annotator agreement and establishing a clear annotation guideline, the remaining reviews were divided into two segments for annotation.

[2]https://shopee.vn/
[3]https://www.lazada.vn/

Aspect	Mean
MATERIAL	Evaluations of the product's materials and fabrics.
DESIGN	The review refers to the style and design of the clothing, e.g., color, shape, feeling of wearing, etc.
PRICE	The review discusses clothing prices and affordability.
SERVICE	The comment mentions sales service, warranty, and delivery.
GENERAL	The review of customers is generally about the product.

Table 1: Aspect definition.

3.3 Statics

The dataset comprises 7,000 reviews, encompassing evaluations across five distinct sentiment aspects. Table 2 presents some samples from our dataset along with their respective aspects and sentiment classifications.

Figure 1 depicts the distribution of each aspect and sentiment within the dataset. Across all aspects, Positive sentiment predominates. Furthermore, over 4,500 reviews are focused on the MATERIAL aspect, comprising more than 60% of all reviews. This highlights the considerable importance customers place on this aspect when making clothing purchases.

The dataset has been thoughtfully partitioned into three distinct sets: 5,000 reviews designated for training, 1,000 reviews for development, and another 1,000 reviews intended for testing. Table 3 presents an overview of the statistics for our dataset.

4 Aspect-based Sentiment Analysis models

The problem is defined as follows, given a review $R = \{S_1, S_2, ..., S_n\}$ with n sentences. The goal is to extract sets of aspects and their corresponding sentiment polarity pairs: $[A_i, SP_i] = LM(R)$. LM denotes the Language Model. The aspect-sentiment polarity pair $[A_i, SP_i] = \{(a_i^k, sp_i^k); a_i^k \in A_i, sp_i^k \in SP_i\}, A = \{a_1, a_2, ..., a_m\}$ is the set of aspects, and $SP = \{sp_1, sp_2, ..., sp_m\}$ is the set of sentiment polarity, with $sp_i \in [positive, negative, neutral]$.

Various methods address the ABSA problem, including rule-based methods (Poria et al., 2014), semantic similarities (Liu et al., 2016), SVM-based algorithms (Jihan et al., 2017), and conditional random fields (CRF) (Shu et al., 2017). Recently, deep neural networks with long short-term memory (LSTM) layers have excelled in extracting sentiment information from word embeddings (Zhang et al., 2018). However, pre-trained language models significantly outperform these methods (Do et al., 2019; Scaria et al., 2023).

Recognizing the potential of language models and the limitations of deep learning models like LSTM, BiLSTM, and GRU for Vietnamese ABSA (Thanh et al., 2021; Mai and Le, 2018), we applied state-of-the-art models using pre-trained language models to our dataset. To ensure compatibility with Vietnamese, we used ViT5, a model pre-trained on Vietnamese (Phan et al., 2022).

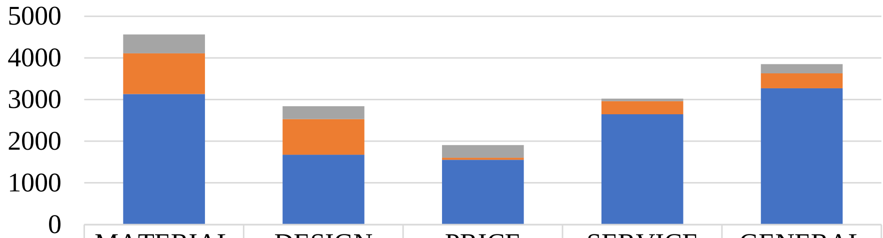
4.1 InstructABSA

From the success of instruction learning (Mishra et al., 2022; Wei et al., 2022), there has been a substantial improvement in the reasoning capabilities of large language models, showcasing impressive results across a variety of tasks. Based on the research by Scaria et al. (2023) we introduce two instruction prompts tailored to the ABSA task. We employ two prompts to facilitate performance comparison, where prompt 1 is translated into Vietnamese from the prompt used by Scaria et al. (2023). Meanwhile, prompt 2 is our proposed prompt. Our approach involves defining these instruction prompts in a manner inspired by the structure depicted in Table 4.

For instruction prompt 1, in addition to the definition, it requires corresponding examples for each sentiment: Positive Example, Neutral Example, and Negative Example. Recognizing that this prompt is relatively lengthy and may increase training time, we proposed prompt 2, which only requests one example that can encompass multiple sentiments. Experimental results indicate that our prompt is higher than the one used by Scaria et al. (2023). The language model LM is refined through instruction tuning using data equipped with instructions, resulting in the instruction-tuned model LM_{Inst}. Subsequently, LM_{Inst} undergoes further fine-tuning for downstream tasks related to ABSA. The task is formulated as follows: $[A_i, SP_i] =$

Review	Aspect & Polarity
Giao hàng nhanh. Nhận được áo đẹp hơn cả mong đợi! Vải áo và đường may rất đẹp, đóng gói rất xịn xò, đánh giá 5 sao. Lần sau sẽ mua ủng hộ shop tiếp ạ. *(Fast delivery. Received a shirt even more beautiful than expected! The fabric and stitching are excellent, and the packaging is very fancy. rated 5 stars. Will support the shop again in the future.)*	SERVICE:positive GENERAL:positive MATERIAL:positive DESIGN:positive
Hàng giao hơi lâu, chất đẹp so với giá tiền rất đáng nhưng màu xanh pastel ở ngoài đậm hơn nhiều so với hình ảnh nên hơi thất vọng một chút. *(The delivery took a bit long, the quality is quite good for the price, but the pastel green color is much darker in person compared to the photo, so I'm a bit disappointed.)*	SERVICE:negative MATERIAL:positive PRICE:positive DESIGN:negative

Table 2: Some samples from the ViCloABSA dataset.



	MATERIAL	DESIGN	PRICE	SERVICE	GENERAL
neutral	449	304	307	60	219
negative	976	859	48	314	357
positive	3135	1675	1555	2646	3273

Figure 1: Distribution of Aspects and Sentiments in the ViCloABSA dataset.

$LM_{Inst}(Inst, R)$.

4.2 MVP

Gou et al. (2023) noted that previous studies often ordered sentiment elements left-to-right, ignoring contrast and language diversity in emotional expression, leading to errors and instability. To address this, they proposed Multi-view Prompting (MVP), which synthesizes predicted emotional factors in various orders. MVP, inspired by prompt chaining (Liu et al., 2021; Wei et al., 2022), leverages different perspectives in human reasoning to control the sequence of emotional elements, enhancing diversity in target expressions.

4.3 GAS

Building on recent successes in framing language tasks as content generation tasks (Raffel et al., 2020; Athiwaratkun et al., 2020; Zhang et al., 2021), we propose addressing ABSA issues with a model that encodes natural language labels into the output. This unified model adapts to multiple tasks without needing task-specific designs.

To facilitate Generative Aspect-based Sentiment Analysis (GAS), we have devised two customized approaches: *GAS-Annotaion* and *GAS-Extraction* modeling. These paradigms reframe the original task as a generation problem. In the former, annotations with label information are added to construct the target sentence. In the latter, the desired natural language label is used directly as the target. The original and target sentences are paired for model training. Additionally, a prediction normalization strategy addresses deviations of generated sentiment elements from the label vocabulary set.

5 Results and discussion

The experimental results on the ViCloABSA dataset for aspect-based sentiment analysis have provided significant insights into the performance of the evaluated methods. Table 5 illustrates the variation of three key metrics: Precision (P), Recall

Set	Review	Positive	Negative	Neutral	Total sentiment
Train	5000	8764	1823	954	11541
Test	1000	1721	405	198	2324
Dev	1000	1799	326	187	2312

Table 3: Statistics of our dataset.

Instruction 1	Definition	Kết quả đầu ra sẽ bao gồm các khía cạnh và cảm xúc của các khía cạnh. Trong trường hợp không có bất kỳ khía cạnh nào, kết quả đầu ra sẽ là "noaspectterm:none". *(The output results will include both aspects and the corresponding emotions for those aspects. In cases where there are no aspects identified, the output result will be "noaspectterm:none.")*
	Example	Input: giao hàng nhanh. nhận đc áo đẹp hơn cả mong đợi! vải áo và đường may rất đẹp đóng gói rất xịn xò, đánh giá 5 sao lần sau sẽ mua ủng hộ shop tiếp ạ. Output: giao hàng nhanh:positive [SEP] đường may rất đẹp:positive *(Input: Fast delivery. The shirt received is more beautiful than expected! The fabric and stitching are excellent, and the packaging is very fancy. I rated it five stars. I will support the shop again in the future.)*
Instruction 2	Definition	Hãy trích xuất ra các khía cạnh và phân loại cảm xúc của các khía cạnh đó. *(Extracting aspects and classifying the corresponding emotions associated with those aspects.)*
	Example	Input: mua size M nhưng cảm thấy hơi bé một xíu, vải mát, đóng gói kĩ, đẹp. Output: vải mát:positive [SEP] hơi bé một xíu:negative [SEP] đóng gói kĩ, đẹp:positive *(Input: bought size M but felt a bit small, cool fabric, carefully packaged, beautiful.)*

Table 4: Instruction prompts.

(R), and F1-score (F) for each evaluation method when using different percentages of the dataset. MVP consistently demonstrates adaptability across different percentages of the dataset, showcasing its robustness in handling varying amounts of training data. For instance, at 5%, MVP achieves a Precision of 34.17, Recall of 31.88, and F1-score of 32.99, while at 100%, these metrics improve to 54.90, 55.94, and 52.61, respectively.

Method	Metrics		
	P	R	F1
MVP	54.90	55.94	52.61
InstructABSA1	**74.00**	**72.12**	**73.11**
InstructABSA2	**74.11**	**72.68**	**73.39**
GAS-Annotation	53.74	51.94	52.83
GAS-Extraction	45.30	44.13	44.71

Table 5: Performance analysis of evaluated methods on ViCloABSA Dataset.

InstructABSA1 and InstructABSA2, two methods utilizing guidance during training, exhibit high performance even with small percentages of the dataset. InstructABSA2 appears more effective, with a substantial increase at higher percentages. At 5%, its Precision, Recall, and F1-score are 59.79, 58.08, and 58.92, respectively, and these values increase to 74.11, 72.68, and 73.39 at 100%.

GAS-Annotation stands out for its Precision, which progressively improves with a larger dataset. However, a corresponding reduction in Recall at higher percentages suggests a potential bias or selectiveness in attention. For example, at 5%, GAS-Annotation achieves a Precision of 9.28 and Recall of 8.52, while at 100%, these metrics change to 53.74 and 51.94, respectively.

GAS-Extraction, while displaying strong Precision, experiences a substantial decline in Recall, emphasizing the delicate balance between these metrics and shedding light on the impact of the chosen extraction methodology. At 5%, GAS-Extraction's Precision, Recall, and F1-score are

38.80, 36.66, and 37.70, and at 100%, these metrics decrease to 45.30, 44.13, and 44.71, respectively.

The performance of these methods with various data segmentations is shown in Figure 2. The trend shows two important points. First, all strong models exhibit an increasing trend in F1-scores as the number of samples in the dataset increases. It indicates that the models can learn and predict more accurately with more data. Second, both InstructABSA1 and InstructABSA2 exhibit high F1-scores, demonstrating robust performance, particularly at higher percentages of data.

6 Conclusion

In the context of advancing research in aspect-based sentiment analysis, this paper introduces ViCloABSA, a meticulously curated dataset designed to propel the field forward. Comprising a substantial collection of over 7,000 human-annotated comments sourced from the domain of clothes e-commerce, ViCloABSA offers a nuanced perspective on sentiment expressions. Each feedback entry undergoes detailed manual annotation, precisely identifying spans relevant to five fine-grained aspect categories, accompanied by their associated sentiment polarities. This study contributes in two major ways. First, the study introduces a specialized Vietnamese dataset centered on clothing reviews from e-commerce platforms, specifically crafted for ABSA tasks. Second, a comprehensive assessment of robust baseline models customized for ABSA tasks is carried out by the research.

We believe that our published dataset will be a valuable resource for future research, promoting further exploration in the field of e-commerce customer feedback analysis. The significant effort invested in ViCloABSA's creation aims to not only provide a comprehensive dataset but also to serve as a catalyst for the development of cutting-edge NLP models.

References

Ben Athiwaratkun, Cicero Nogueira dos Santos, Jason Krone, and Bing Xiang. 2020. Augmented natural language for generative sequence labeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 375–385, Online. Association for Computational Linguistics.

Jiahao Bu, Lei Ren, Shuang Zheng, Yang Yang, Jingang Wang, Fuzheng Zhang, and Wei Wu. 2021. Asap: A chinese review dataset towards aspect category sentiment analysis and rating prediction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, page 2069–2079.

Hai Ha Do, Angelika Maag, and Abeer Alsadoon. 2019. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert systems with applications*, 118:272–299.

Zhibin Gou, Qingyan Guo, and Yujiu Yang. 2023. MvP: Multi-view prompting improves aspect sentiment tuple prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4380–4397, Toronto, Canada. 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. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 6280–6285.

Nadheesh Jihan, Yasas Senarath, Dulanjaya Tennekoon, Mithila Wickramarathne, and Surangika Ranathunga. 2017. Multi-domain aspect extraction using support vector machines. In *Proceedings of the 29th conference on computational linguistics and speech processing (ROCLING 2017)*, pages 308–322.

Bing Liu. 2020. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55:1 – 35.

Qian Liu, Bing Liu, Yuanlin Zhang, Doo Soon Kim, and Zhiqiang Gao. 2016. Improving opinion aspect extraction using semantic similarity and aspect associations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Luong Luc Phan, Phuc Huynh Pham, Kim Thi-Thanh Nguyen, Sieu Khai Huynh, Tham Thi Nguyen, Luan Thanh Nguyen, Tin Van Huynh, and Kiet Van Nguyen. 2021. Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence. In *Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part II 14*, pages 647–658. Springer.

Long Mai and Bac Le. 2018. Aspect-based sentiment analysis of vietnamese texts with deep learning. In *Intelligent Information and Database Systems: 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings, Part I 10*, pages 149–158. Springer.

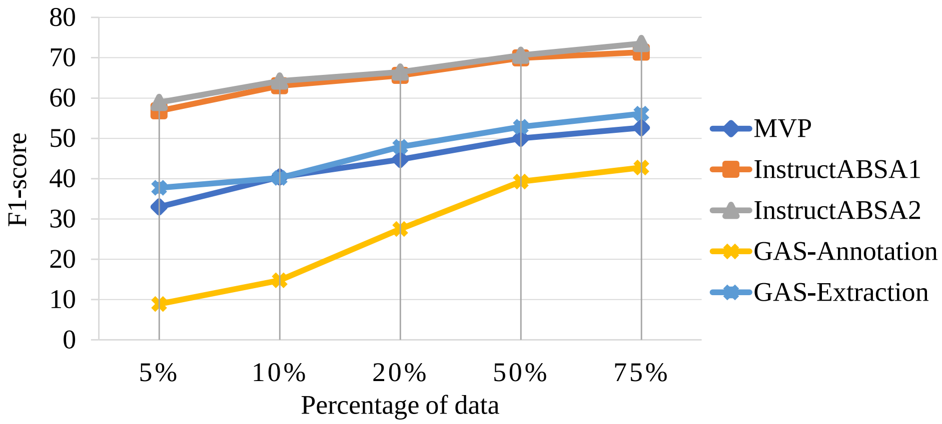


Figure 2: Model performance across different percentages of the ViCloABSA dataset.

Swaroop Mishra, Daniel Khashabi, Chitta Baral, and Hannaneh Hajishirzi. 2022. Cross-task generalization via natural language crowdsourcing instructions. In *60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*, pages 3470–3487. Association for Computational Linguistics (ACL).

Huyen TM Nguyen, Hung V Nguyen, Quyen T Ngo, Luong X Vu, Vu Mai Tran, Bach X Ngo, and Cuong A Le. 2018. Vlsp shared task: sentiment analysis. *Journal of Computer Science and Cybernetics*, 34(4):295–310.

Long Phan, Hieu Tran, Hieu Nguyen, and Trieu H Trinh. 2022. Vit5: Pretrained text-to-text transformer for vietnamese language generation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, pages 136–142.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Soujanya Poria, Erik Cambria, Lun-Wei Ku, Chen Gui, and Alexander Gelbukh. 2014. A rule-based approach to aspect extraction from product reviews. In *Proceedings of the second workshop on natural language processing for social media (SocialNLP)*, pages 28–37.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. 2016. Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1546–1556.

Kevin Scaria, Himanshu Gupta, Siddharth Goyal, Saurabh Arjun Sawant, Swaroop Mishra, and Chitta Baral. 2023. Instructabsa: Instruction learning for aspect based sentiment analysis. *arXiv preprint arXiv:2302.08624*.

Lei Shu, Hu Xu, and Bing Liu. 2017. Lifelong learning crf for supervised aspect extraction. *arXiv preprint arXiv:1705.00251*.

Kim Nguyen Thi Thanh, Sieu Huynh Khai, Phuc Pham Huynh, Luong Phan Luc, Duc-Vu Nguyen, and Kiet Nguyen Van. 2021. Span detection for aspect-based sentiment analysis in vietnamese. In *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, pages 318–328.

Quang-Linh Tran, Phan Thanh Dat Le, and Trong-Hop Do. 2022. Aspect-based sentiment analysis for Vietnamese reviews about beauty product on E-commerce websites. In *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, pages 767–776, Manila, Philippines. Association for Computational Linguistics.

Kiet Van Nguyen, Vu Duc Nguyen, Phu XV Nguyen, Tham TH Truong, and Ngan Luu-Thuy Nguyen. 2018. Uit-vsfc: Vietnamese students' feedback corpus for sentiment analysis. In *2018 10th international conference on knowledge and systems engineering (KSE)*, pages 19–24. IEEE.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Ting Xu, Huiyun Yang, Zhen Wu, Jiaze Chen, Fei Zhao, and Xinyu Dai. 2023. Measuring your ASTE models in the wild: A diversified multi-domain dataset for aspect sentiment triplet extraction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 2837–2853, Toronto, Canada. Association for Computational Linguistics.

Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.