Comparative Analysis of Recent Studies on Aspect-Based Sentiment Analysis

Faiz Ghifari Haznitrama, Ho-Jin Choi
School of Computing, KAIST

haznitrama@kaist.ac.kr, hojinc@kaist.ac.kr

Abstract

Sentiment analysis as part of natural language processing (NLP) has received much attention following the demand to understand people's opinions. Aspect-based sentiment analysis (ABSA) is a fine-grained subtask from sentiment analysis that aims to classify sentiment at the aspect level. Throughout the years, researchers have formulated ABSA into various tasks for different scenarios. Unlike the early works, the current ABSA utilizes many elements to improve performance and provide more details to produce informative results. These ABSA formulations have provided greater challenges for researchers. However, it is difficult to explore ABSA's works due to the many different formulations, terms, and results. In this paper, we conduct a comparative analysis of recent studies on ABSA. We mention some key elements, problem formulations, and datasets currently utilized by most ABSA communities. Also, we conduct a short review of the latest papers to find the current state-of-the-art model. From our observations, we found that span-level representation is an important feature in solving the ABSA problem, while multi-task learning and generative approach look promising. Finally, we review some open challenges and further directions for ABSA research in the future.

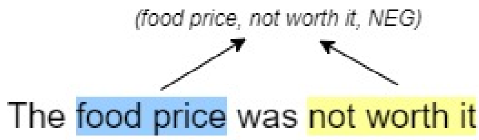
1. Introduction

Aspect-based sentiment analysis (ABSA) has gained significant attention in the past decade as a subtask of sentiment analysis that predicts sentiment for each aspect of a text, providing more informative predictions compared to conventional sentiment analysis. Initially, sentiment analysis was limited to predicting the general sentiment polarity of a text or document without considering different aspects. However, ABSA introduces aspects to the task, increasing the complexity and requiring the model to understand the context in more detail. ABSA datasets have grown to cover more detailed features, and researchers have reformulated their approach to solving the ABSA problem, resulting in more complexity and a better understanding of predictions, which poses a challenge for the model to comprehend the appropriate context of a text and its features.

In this paper, we compare recent ABSA research and analyze the approaches and methods proposed, focusing on studies that formulate ABSA to provide more information on predicted sentiment. Our criteria for included papers are published in 2019 and above, using most ABSA elements as features, and the ability to predict more than just sentiment and aspect. We also discuss potential challenges, opportunities, and future directions for ABSA.

2. Aspect-Based Sentiment Analysis

These are some key elements that are mostly used in the latest ABSA research. Consider an example as shown in Figure 1 “*The food price was not worth it*”. Here, we can have an aspect *a* “*food price*”, a category *c* “*finance*”, an opinion *o* “*not worth it*”, and a sentiment *s* “*negative*”. However, this depends on the approach and problem formulation used by the researchers. In the sample shown in Figure 1, it used Aspect Sentiment Triplet Extraction (ASTE) [1] task as formulation. Thus, the prediction will be in a triplet format of (aspect, opinion, sentiment). In this paper, we cover only the ASTE and Aspect Category Opinion Sentiment (ACOS) Quadruplet Extraction [2] tasks.



(Figure 1) An example of ABSA formulated with triplet extraction task. The blue text denotes the aspect term, and the yellow text denotes the opinion term.

Effective ABSA models require considering various factors such as aspects, targets, opinions, domain, learning approach, and feature representation. While supervised learning is mostly utilized, unsupervised, reinforcement, and multi-task learning approaches are becoming more feasible. Pretrained language models have become more popular, and domain adaptation techniques can transfer knowledge from source to target domains to improve model performance. Text features can be represented using various methods, including encoder representation and label representation.

3. Frameworks and Models

There are two main frameworks used in aspect-based sentiment analysis (ABSA): the end-to-end framework and

<Table 1> Evaluation results on ASTE-Data-V2 [3]. *P.*, *R.*, and F_1 mean precision, recall, and F_1-score respectively.
- means there is no information about that specific score from the paper.

Models	**Dataset**											
	14Res			14Lap			15Res			16Res		
	P.	*R.*	F_1.	*P.*	*R.*	F_1.	*P.*	*R.*	F_1.	*P.*	*R.*	F_1.
End-to-End												
JET [3]	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	70.42	58.37	63.83
GTS [4]	68.71	67.67	68.17	58.54	50.65	54.30	60.69	**60.54**	60.61	67.39	66.73	67.06
GAS [5]	-	-	72.16	-	-	60.78	-	-	62.10	-	-	70.10
SBN [6]	**76.36**	**72.43**	**74.34**	**65.68**	**59.88**	**62.65**	**69.93**	60.41	**64.82**	**71.59**	**72.57**	**72.08**
Extract-then-Classify												
ASTE-RL [8]	69.70	69.23	69.47	64.80	54.99	59.50	63.31	61.61	62.44	64.76	70.74	67.57
CEN [9]	63.59	**73.44**	68.16	57.84	**59.33**	58.58	54.53	63.30	58.59	63.57	**71.98**	67.52
Span-ASTE [10]	**72.89**	70.89	71.85	63.44	55.84	59.38	62.18	**64.45**	**63.27**	**69.45**	71.17	70.26
B-MRC [11]	71.32	70.09	70.69	**65.12**	54.41	59.27	**63.71**	58.63	61.05	67.74	68.56	68.13
NAT [12]	-	-	71.37	-	-	55.07	-	-	61.83	-	-	70.06
SyMux [13]	-	-	**74.84**	-	-	**60.11**	-	-	63.13	-	-	**72.76**

the extract-then-classify framework. The end-to-end framework combines the extractor and classifier models into one, resulting in a more complex model but potentially improving performance. Several models utilize the end-to-end framework. **JET** [3] uses a new position-aware tagging scheme and combines an encoder with CRF for the tag scoring function. **GTS** [4] is an end-to-end approach to ABSA that utilizes an upper triangular grid schema, with an encoder along with custom inference and decoding strategies. **GAS** [5] is a generative model introduced that transforms ABSA into a text generation task. **SBN** [6] is a span-level ABSA model that employs bidirectional networks from both aspect and opinion decoders to predicts sentiment directly, Last is **BART-ABSA** [7], a modified version of the GAS model to handle implicit aspects and opinions.

On the other hand, the extract-then-classify framework involves breaking down the task into smaller extraction and classification tasks, each handled by a separate model. Several models have also employed this framework. **ASTE-RL** [8] proposes a reinforcement learning-based model that involves sentiment classification that triggers the extraction part. **CEN** [9] introduce perceivable pairs as supervision. **Span-ASTE** [10] utilized a span-level representation to capture the interactions between aspects and opinions. **B-MRC** [11] utilizes a bidirectional MRC approach with non-restrictive and restrictive queries for sentiment prediction. **NAT** [12] utilizes non-autoregressive approach to capture overlapping structures between aspect, opinion, and sentiment terms. **SyMux** [13] unifies seven subtasks through a multiplex cascade framework, incrementally decoding the small subtasks until reaching high-level task. **EC-ACOS** [2] model utilizes aspect and opinion extraction and category-sentiment classification of aspect-opinion pairs to detect implicit aspect or opinion.

4. Analysis

We compare and analyze the results of the papers explained above to determine which work performs better

<Table 2> Evaluation results on ACOS [2]. *P.*, *R.*, and F_1 mean precision, recall, and F_1-score respectively.

Models	**Dataset**					
	Restaurant-ACOS			Laptop-ACOS		
	P.	*R.*	F_1.	*P.*	*R.*	F_1.
JET-ACOS [3]	**59.8**	28.9	39.0	44.5	16.2	23.8
EC-ACOS [2]	38.5	**52.9**	44.6	**45.5**	29.4	35.8
BART-ABSA [7]	56.8	51.1	**53.5**	41.1	**37.9**	**39.4**

and identify factors that affect model performance. We evaluate two sets of results: those from papers that benchmarked on ASTE using ASTE-Data-V2 [3], and those from the new ACOS quadruplet extraction task.

Table 1 shows the evaluation results of all papers on ASTE-Data-V2, where SBN is the best model overall, while SyMux performs best on specific datasets. However, model with the highest F_1-score did not necessarily have the best precision or recall. The CEN model has the best recall on 14Res, and JET has a high precision score but low recall. GAS, Spas-ASTE, B-MRC, and NAT are the other models that perform well. Although they all provide a similar number of F_1-score with a margin of 1-2%, they have different characteristics, such as Span-ASTE being strong on recall while B-MRC is ahead on precision. ASTE-RL and CEN also produce relatively good performance, while JET falls short.

We hypothesized that the end-to-end framework would be ideal. The evaluation results in Table 1 showed that our hypothesis is not necessarily wrong, as the SBN model successfully performed as the best model while SyMux also had a strong performance. The importance of span-level features in ABSA models was also highlighted, with the SBN and Span-ASTE model utilizing span-level interactions. SyMux cascading multi-task learning framework was also notable, showing that multi-task learning can be beneficial for ABSA with the right choice. The ability to reduce model complexity in terms of filtering invalid triplets and the importance of interactions between aspects and opinions were also noted, with SBN and B-MRC models both utilizing a bidirectional approach and SyMux enhancing features by producing a 2-D representation.

Table 2 shows the evaluation results for three models on the ACOS dataset. The EC-ACOS model outperforms JET-ACOS, but BART-ABSA is the best overall model to handle both explicit and implicit aspect-opinion terms. EC-ACOS can handle implicit elements but suffers from a lower F1 score than JET-ACOS in explicit elements.

5. Future Work

The future of ABSA research is expected to focus on the end-to-end framework, span-level feature extraction, generative models, and open-domain ABSA. The end-to-end framework offers a performance gain and opportunities for improvement, while the success of SBN and Span-ASTE highlights the importance of span-level features. The use of generative models and text-to-text format could be a new trend in ABSA, while open-domain ABSA is needed to meet the demands public usage. More challenging open-domain ABSA datasets and models are expected to emerge in the future.

6. Conclusion

We provide a comparative analysis of recent studies on ABSA, highlighting two popular frameworks: end-to-end and extract-then-classify. The extract-then-classify model SyMux achieves competitive results with the best end-to-end model SBN due to its use of cascading multi-task learning. The SBN model uses span-level interactions and a bidirectional aspect-opinion decoder to achieve great results. The ACOS task is also interesting, with the EC-ACOS and BART-ABSA models performing well on implicit and explicit features. We suggest that more research is needed on the ACOS benchmark and open-domain ABSA.

Acknowledgements

This work was supported by the Technology Innovation Program (Grant number: 20012288, Development of Affective Virtual TA service based on deep learning for foreign language education) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea)

References

[1] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis", *AAAI*, vol. 34, no. 05, pp. 8600-8607, Apr. 2020.

[2] H. Cai, R. Xia, and J. Yu, "Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions", 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)*, 2021, pp. 340–350.

[3] L. Xu, H. Li, W. Lu, and L. Bing, "Position-Aware Tagging for Aspect Sentiment Triplet Extraction", in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 2339–2349.

[4] Z. Wu, C. Ying, F. Zhao, Z. Fan, X. Dai, and R. Xia, "Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction", in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 2576–2585.

[5] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "Towards generative aspect-based sentiment analysis", in *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)*, 2021, pp. 504–510.

[6] Y. Chen, C. Keming, X. Sun, and Z. Zhang, "A Span-level Bidirectional Network for Aspect Sentiment Triplet Extraction", in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022, pp. 4300–4309.

[7] C. D. Hoang, Q. V. Dinh, and N. H. Tran, "Aspect-Category-Opinion-Sentiment Extraction Using Generative Transformer Model", in *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2022, pp. 1–6.

[8] S. Yu Bai Jian, T. Nayak, N. Majumder, and S. Poria, "Aspect sentiment triplet extraction using reinforcement learning", in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 3603–3607.

[9] L. Huang et al., "First Target and Opinion then Polarity: Enhancing Target-opinion Correlation for Aspect Sentiment Triplet Extraction", *arXiv [cs.CL]*. 2021.

[10] L. Xu, Y. K. Chia, and L. Bing, "Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction", 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)*, 2021, pp. 4755–4766.

[11] S. Chen, Y. Wang, J. Liu, and Y. Wang, "Bidirectional machine reading comprehension for aspect sentiment triplet extraction", in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, pp. 12666–12674.

[12] H. Fei, Y. Ren, Y. Zhang, and D. Ji, "Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction", *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[13] H. Fei, F. Li, C. Li, S. Wu, J. Li, and D. Ji, "Inheriting the wisdom of predecessors: A multiplex cascade framework for unified aspect-based sentiment analysis", in *Proc. 31st International Joint Conference on Artificial Intelligence*, 2022, pp. 4096–4103.