ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095101

SENER: SENTIMENT ELEMENT NAMED ENTITY RECOGNITION FOR ASPECT-BASED SENTIMENT ANALYSIS

*Sun-Kyung Lee, Jong-Hwan Kim**

KAIST, Daejeon, Republic of Korea

ABSTRACT

Aspect-based sentiment analysis (ABSA) is a task of identifying fine-grained sentiment entities in a given sentence, which is generally formulated as a sequence labeling problem. Recently, advancements in large pre-trained language models (PLMs) led to generative ABSA, where the task is treated as text-to-text transition resolved by fine-tuning PLMs. Although the generative methods are designed to capture sentence-level semantic information, they are inappropriate for explicit comprehension of sentiment structure. In order to address this issue, we propose sentiment element named entity recognition (SENER) for ABSA. SENER integrates the concepts of named entity recognition (NER) and generative ABSA to retrieve the sentiment entities with predefined sentiment element names, leading to better semantic and sentiment structure understanding. Experimental results on several ABSA tasks show that the proposed SENER significantly outperforms previous works on ASQP and ASTE.

Index Terms— Aspect-based Sentiment Analysis, Named Entity Recognition, Sequence Labeling Problem, Pre-trained Language Models

1. INTRODUCTION

Aspect-based sentiment analysis (ABSA) has long been studied in natural language processing (NLP) for its applicability to real world scenarios. The goal of ABSA is to identify groups of aspect-level sentiment entities, where four fundamental elements are generally considered [1]; aspect term, aspect category, opinion term and sentiment polarity. Given an example of *"The staff is friendly."*, the aspect term is *"staff"*, the category it belongs to is *"service general"*, the opinion term is *"friendly"* and the sentiment towards the aspect is *"positive"*. Depending on the number of sentiment elements to retrieve, there are multiple sub-tasks of ABSA. Earlier works such as aspect term extraction [2, 3] or aspect category detection [4, 5] focus on single sentiment element. More recent works, such as aspect sentiment triplet extraction (ASTE) [6, 7, 8], target aspect sentiment detection (TASD) [9, 10, 11] and aspect sentiment quad prediction (ASQP) [12], handle three or all four sentiment elements.

*Corresponding Author

Input Sentence: The staff is friendly.

Method	Target Sentence
GAS-A	The [staff | friendly | service general | positive] is friendly.
GAS-E	(staff, friendly, service general, positive)
Paraphrase	service general is great because staff is friendly
Opinion Tree	(Root, (Quad, (service general, staff), (positive, friendly)))
SENER	aspect is **staff**, opinion is **friendly**, category is **service general**, sentiment is **positive**

Fig. 1. Examples of the generated target sentences depending on different approaches.

Previously, ABSA is usually formulated as a sequence labeling problem by performing sequence-level classification. While this approach could explicitly designate each sequence to corresponding sentiment element, it easily loses semantic information as it treats the label as number indices. More recently, text-to-text generation approach is gaining interest with the success of large pre-trained language models (PLMs). Given an input sentence, PLMs are fine-tuned to generate the target sentence containing the sentiment entities as shown in Fig. 1. For example, GAS [13] used either annotation- (GAS-A) or extraction-style (GAS-E) to represent the target sentence. [12] presented paraphrase paradigm, which can be generalized to various ABSA tasks. Also, [14] introduced opinion tree generation to model semantic representation. After the generation, the sentiment entities can be recovered from the target sentence.

Even though these generative methods are capable of capturing strong semantic relations among sentiment entities, they only have implicit understanding of the exact role of each entity as sentiment element. That is, there is no explicit indication in the generated target sentence whether "staff" is the aspect term or "friendly" is the opinion term, thus leading to poor sentiment structure comprehension. To tackle this issue, we bring the concept of named entity recognition (NER) into ABSA and propose sentiment element named entity recognition (SENER).

NER [15] is a task of extracting entities from a given text into pre-defined names. Similarly, specific names for each sentiment element (such as "aspect", "opinion", "category" and "sentiment" in Fig. 1) is defined in SENER. SENER is designed to capture sentiment entities with corresponding sentiment element names. This way, SENER can better understand the sentiment structure. Also, by maintaining the generative ABSA framework for training and inferring, SENER is able to exploit rich semantic information from the PLMs. Evaluated on various ABSA tasks, the proposed SENER shows the state-of-the-art results on ASQP and ASTE and comparable results on TASD. In short, the main contributions of this paper are summarized as follows:

- We effectively formulate and integrate the two important research areas in NLP; NER and ABSA
- We propose SENER, which successfully captures both sentiment structure and semantic information.
- SENER is verified on various ABSA tasks with extensive experiments.

2. METHODOLOGY

Here, we first formulate the task of NER and ABSA separately. Note that ASQP (retrieving all four sentiment elements at once) is considered as the most generalized form of ABSA. Then the concepts of NER and ABSA are integrated as sentiment element named entity recognition (SENER).

2.1. Problem Formulation

The task of NER [16] can be formulated as recognizing all entity occurrences Y_N given a sentence X. Suppose V_X as a set containing all possible continuous spans from X, the i-th entity occurrence $y_{N_i} \in Y_N$ is defined as $y_{N_i} = (v_i, e_i)$, where v_i is a span $v_i \in V_X$. And $e_i \in E$ is an entity name, which is the element from the full set of entity names E.

On the other hand, the task of ASQP [12] can be formulated as recognizing all sentiment entity tuples Y_A given a sentence X. The i-th sentiment entity tuple occurrence $y_{A_i} \in Y_A$ is defined as $y_{A_i} = (a_i, o_i, c_i, s_i)$, where each corresponds to the aspect term, opinion term, aspect category, and sentiment polarity. Considering the cases where the aspect term and opinion term are not explicitly mentioned in X, $a_i, o_i \in V_X \cup \{NULL\}$. The aspect category c_i is included in the full set of categories C and the sentiment polarity is either positive, neutral or negative, that is $s_i \in \{positive, neutral, negative\}$.

2.2. Sentiment Element Named Entity Recognition

Combining the formulation of NER and ASQP, the task of SENER can be formulated as recognizing all the sentiment

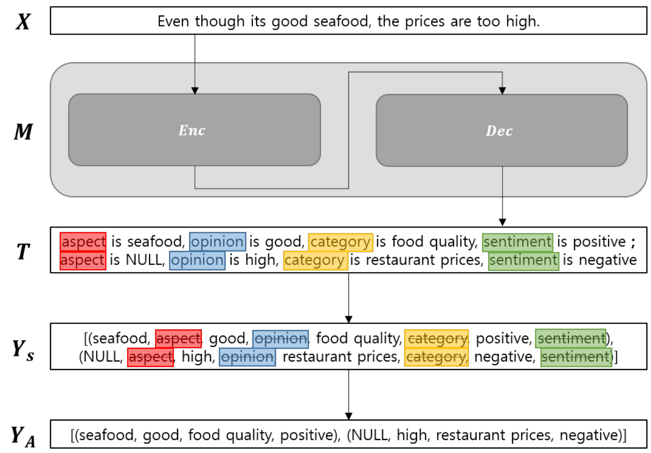

Fig. 2. Overall flow of SENER.

entities along with corresponding names. Formally, recognize Y_S given X, where the i-th tuple occurrence $y_{S_i} \in Y_S$ is defined as $y_{S_i} = (a_i, e_a, o_i, e_o, c_i, e_c, s_i, e_s)$. $e_a, e_o, e_c, e_s \in E$ and indicate the fixed sentiment element name for the aspect term, opinion term, aspect category, and sentiment polarity, respectively. The sentiment element names can be either phrases composed of the tokens from the original vocabulary of PLMs or specific synthetic tokens which is newly added to the vocabulary. Note that SENER can be easily generalized to other ABSA tasks. As an example, y_{S_i} can be $(a_i, e_a, o_i, e_o, s_i, e_s)$ for ASTE (retrieving the aspect term, the opinion term and the sentiment polarity) and similarly, $(a_i, e_a, c_i, e_c, s_i, e_s)$ for TASD (retrieving the aspect term, the aspect category and the sentiment polarity).

2.3. SENER as Text-to-Text Generation

SENER maintains the generative ABSA framework to exploit semantic information from sentiment entities and sentiment element names. To ease the sequence-to-sequence learning, Y_S is first transformed to target sentence in natural languages $T = (t_1, ..., t_{|T|})$ with sequence length $|T|$, as shown in Fig. 2. Simply, i-th tuple y_{S_i} can be linearized as follows:

"e_a **is** a_i, e_o **is** o_i, e_c **is** c_i, e_s **is** s_i"

where "is" indicates that the sentiment entity corresponds to the sentiment element name and "," indicates that the sentiment entities are correlated. As there can be multiple sentiment entity tuples in a given sentence, ";" is used to separate linearized tuples.

For the sequence-to-sequence learning, encoder-decoder architecture such as transformer [17] can be utilized. Suppose a model M composed of encoder Enc and decoder Dec parameterized by θ. Following [18], the goal of M is to estimate the conditional probability of the target sentence given an input sentence, which is $p(T|X)$. With $H = Enc(X)$ as the

	ASQP		ASTE				TASD	
	Rest15	Rest16	Lap14	Rest14	Rest15	Rest16	Rest15	Rest16
Train	834	1264	906	1266	605	857	1120	1708
Dev	209	316	219	310	148	210	-	-
Test	537	544	328	492	322	326	582	587

Table 1. Number of Train/Dev/Test sentences in the datasets used for the experiments.

contextualized hidden representation of X by the encoder, the conditional probability can be rewritten as follows:

$$p(T|X) = \prod_{i=1}^{|T|} p(t_i|T_{<i}, H)$$

where $p(t_i|T_{<i}, H)$ is the softmax normalized probability over the vocabulary of PLMs and defined as:

$$p(t_i|T_{<i}, H) = \begin{cases} p(t_1|H) & \text{if } i\text{=1} \\ p(t_i|t_1, ..., t_{i-1}, H) & \text{if } i\text{>1} \end{cases}$$

2.4. Training and Inference

According to [18], given a train set D with size $|D|$, the sequence-to-sequence model M can be trained by maximizing the log-likelihood of a correct target sentence as follows:

$$\frac{1}{|D|} \sum_{(X,T) \in D} \log p(T|X; \theta).$$

After finishing training, the overall flow of inferring is shown in Fig. 2. The trained model M takes X as an input and returns the conditional probability $p(T|X)$. Then, the greedy decoding is applied to decode the conditional probability into the target sentence T. Specific separators such as "is", "," and ";" are removed from T and Y_S can be generated. Lastly, sentiment element names are eliminated from Y_S for the final result of Y_A.

3. EXPERIMENTS

3.1. Experimental Settings

Three different ABSA tasks are considered for the experiments; ASQP, ASTE and TASD. All datasets originate from the SemEval challenges [19] while the datasets are slightly modified to fit each task. For a fair comparison, datasets from the previous works are adopted with the same splits. There are four domains (Lap14, Rest14, Rest15 and Rest16) in the ASTE dataset [7] and two domains (Rest15 and Rest16) in the ASQP [12] and TASD datasets [9]. The statistics of the datasets are summarized in Table 1.

For evaluation, F1 score is adopted. A sentiment entity tuple inferred by the model is counted as true positive if and only if all the entities are exactly the same as the ground truth labels. All the reported scores are averaged over five different runs. With regard to training details, T5-base model [20] is adopted as the sequence-to-sequence model M. The pre-trained weights are loaded and M is fine-tuned to each ABSA task with a cross-entropy loss. The model is trained on one Nvidia RTX 3090 GPU with the hyperparameters set as follows: batch size is 16, learning rate is 3e-4 with AdamW optimizer [21] and epoch is 20.

3.2. Control Groups

Recent works used as baselines for the experiments are summarized. Most of them take the advantage of large PLMs such as BERT [22] or T5 [20] and the list is as follows:

- GAS [13] (ASQP, ASTE, TASD): A unified generative framework for ABSA with T5. Especially, extraction-style (GAS-EXTRACTION) is used. GAS is originally designed to retrieve less or equal to three sentiment elements, but it is slightly modified in this experiment to handle all four sentiment elements.

- Paraphrase [12] (ASQP, ASTE, TASD): Paraphrases the input sentence to natural language form containing sentiment information with T5.

- Opinion Tree [14] (ASQP): Generates an opinion tree that can model semantic representation with T5. The original paper exploits additional pre-training and constrained decoding. Here, we just take the architecture of the linearized tree and apply greedy decoding to focus on comparing how well a model itself captures the sentiment structure.

- JET [7] (ASTE): Jointly extracts the triplets in an end-to-end manner based on a position-aware tagging scheme. One model variant with BERT contextualized word representation (JET$^{O}_{+BERT}$) is used.

- GTS [23] (ASTE): A grid tagging scheme is introduced for end-to-end ABSA. Similar to JET, BERT based variant (GTS-BERT) is used.

- Two-Stage [24] (ASTE): A two-stage framework which enhances the correlation between target-opinion pairs by BERT based sequence tagging first to detect targets and opinions, and then estimating sentiment.

- TAS [9] (TASD): First to tackle TASD task with various configurations of BERT based models. TASD-BERT-SW-BIO-CRF and TAS-BERT-SW-TO are used.
- MEJD [10] (TASD): End-to-end multiple element joint detection model which employs BERT for initial embedding and models dependency relationship with Bi-LSTM and graph convolutional network.

There are two variants of the proposed SENER depending on whether to use the original or synthetic tokens as sentiment element names (mentioned in Section 2.2). For SENER-orig, e_a, e_o, e_c, e_s are set to be "aspect", "opinion", "category" and "sentiment", respectively, which exist in the pre-trained vocabulary of T5. On the other hand, specific synthetic tokens corresponding to each sentiment element name are newly added to the vocabulary for SENER-syn. The embeddings of these synthetic tokens are randomly initialized and trained along with the model.

3.3. Results and Analysis

Model	Rest15	Rest16
GAS	45.98	56.04
Paraphrase	46.93	57.93
Opinion Tree	<u>47.60</u> $\pm$1.13	58.07 $\pm$0.61
SENER-orig	**48.45** $\pm$0.16	<u>58.46</u> $\pm$1.01
SENER-syn	47.19 $\pm$0.58	**59.40** $\pm$0.47

Table 2. Experimental results on ASQP

Model	Lap14	Rest14	Rest15	Rest16
JET	51.04	62.04	57.53	63.83
GTS	55.21	64.81	54.88	66.08
Two-Stage	58.58	68.16	58.59	67.52
GAS	58.19	70.52	60.23	69.05
Paraphrase	<u>61.13</u>	72.03	62.56	<u>71.70</u>
SENER-orig	**61.94**	<u>72.23</u>	**65.38**	**73.35**
SENER-syn	59.96	**73.78**	<u>63.10</u>	71.37

Table 3. Experimental results on ASTE

Tables 2, 3 and 4 show the experimental results on ASQP, ASTE and TASD, respectively. SENER significantly outperforms other models on ASQP and ASTE, which implies the power of SENER in capturing sentiment structure and semantic information. The experimental results demonstrates the effectiveness of the proposed framework.

Comparing SENER-orig and SENER-syn in Tables 2 and 3, the performance varies depending on the task or the dataset. Note that SENER-orig achieves better F1 scores than previous baselines in every setting, while there are some deviations in SENER-syn. Also, there is a correlation between the performance and the number of training samples in the dataset. That is, SENER-orig performs better when the train set is relatively small, while SENER-syn performs better when the train set is relatively large. This observation becomes clear when referring to the dataset statistics in Table 1. One possible reason is that if there are insufficient resources for training, SENER is dependent on the already known semantic relations. Thus, SENER-orig takes the full advantage of PLMs leading to stable performance, but newly added synthetic tokens in SENER-syn hinder it. On the other hand, when sufficient training data is provided, even the synthetic tokens can learn rich semantic representation. Also, the synthetic tokens can better understand their role as sentiment element names as they are trained from scratch.

Model	Rest15	Rest16
TAS-CRF	57.51	65.89
TAS-TO	58.09	65.44
MEJD	57.76	67.66
GAS	60.63	68.31
Paraphrase	**63.06**	**71.97**
SENER-orig	<u>63.04</u>	69.74
SENER-syn	61.94	<u>71.27</u>

Table 4. Experimental results on TASD

In Table 4, unlike the results on ASQP and ASTE, SENER shows just comparable results on TASD. This is because ABSA tasks are actually the compound of extraction and classification tasks [25]. For each sentiment element, the aspect term and the opinion term are the results of extraction from the sentence, while the aspect category and the sentiment polarity are the results of classification. In this point of view, two extractions and two classifications should be performed in ASQP for one sentiment entity tuple. Similarly, there are two extractions and one classification for ASTE, and one extraction and two classifications for TASD. Note that SENER is inspired by NER which is is originally oriented for information extraction tasks [15]. Therefore, it can be hypothesized that SENER has its weakness in classification dominant ABSA tasks such as TASD. Conversely, SENER enhances the performance in ASQP and ASTE, where extraction takes at least a half portion of the entire tasks.

4. CONCLUSION

In this paper, we proposed SENER for ABSA. Through the integration of NER and generative ABSA, SENER could explicitly model both sentiment structure and semantic information. And extensive experiments on several ABSA tasks demonstrated the SENER formulation. Also, analysis on experimental results revealed that SENER reinforces the extraction ability for ABSA. For further works, we plan to boost the classification ability for ABSA.

5. REFERENCES

[1] Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," *arXiv preprint arXiv:2203.01054*, 2022.

[2] Pengfei Liu, Shafiq Joty, and Helen Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 1433–1443.

[3] Hu Xu, Bing Liu, Lei Shu, and Philip S Yu, "Double embeddings and cnn-based sequence labeling for aspect extraction," *arXiv preprint arXiv:1805.04601*, 2018.

[4] Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao, "Representation learning for aspect category detection in online reviews," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[5] Shiva Ramezani, Razieh Rahimi, and James Allan, "Aspect category detection in product reviews using contextual representation," in *Proceedings of ACM SIGIR Workshop on eCommerce (SIGIR eCom'20)*, 2020.

[6] Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, pp. 8600–8607.

[7] Lu Xu, Hao Li, Wei Lu, and Lidong Bing, "Position-aware tagging for aspect sentiment triplet extraction," *arXiv preprint arXiv:2010.02609*, 2020.

[8] Lu Xu, Yew Ken Chia, and Lidong Bing, "Learning span-level interactions for aspect sentiment triplet extraction," *arXiv preprint arXiv:2107.12214*, 2021.

[9] Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z Pan, "Target-aspect-sentiment joint detection for aspect-based sentiment analysis," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 9122–9129.

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

[11] Cai Ke, Qingyu Xiong, Chao Wu, Zikai Liao, and Hualing Yi, "Prior-bert and multi-task learning for target-aspect-sentiment joint detection," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 7817–7821.

[12] Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam, "Aspect sentiment quad prediction as paraphrase generation," *arXiv preprint arXiv:2110.00796*, 2021.

[13] Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai 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.

[14] Xiaoyi Bao, Wang Zhongqing, Xiaotong Jiang, Rong Xiao, and Shoushan Li, "Aspect-based sentiment analysis with opinion tree generation," .

[15] Vikas Yadav and Steven Bethard, "A survey on recent advances in named entity recognition from deep learning models," *arXiv preprint arXiv:1910.11470*, 2019.

[16] Liwen Wang, Rumei Li, Yang Yan, Yuanmeng Yan, Sirui Wang, Wei Wu, and Weiran Xu, "Instructionner: A multi-task instruction-based generative framework for few-shot ner," *arXiv preprint arXiv:2203.03903*, 2022.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[18] Ilya Sutskever, Oriol Vinyals, and Quoc V Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.

[19] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al., "Semeval-2016 task 5: Aspect based sentiment analysis," in *International workshop on semantic evaluation*, 2016, pp. 19–30.

[20] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al., "Exploring the limits of transfer learning with a unified text-to-text transformer.," *J. Mach. Learn. Res.*, vol. 21, no. 140, pp. 1–67, 2020.

[21] Ilya Loshchilov and Frank Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

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

[23] Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia, "Grid tagging scheme for aspect-oriented fine-grained opinion extraction," *arXiv preprint arXiv:2010.04640*, 2020.

[24] Lianzhe Huang, Peiyi Wang, Sujian Li, Tianyu Liu, Xiaodong Zhang, Zhicong Cheng, Dawei Yin, and Houfeng Wang, "First target and opinion then polarity: Enhancing target-opinion correlation for aspect sentiment triplet extraction," *arXiv preprint arXiv:2102.08549*, 2021.

[25] Hongjie Cai, Rui Xia, and Jianfei 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.