

Towards Generative Aspect-Based Sentiment Analysis*

Wenxuan Zhang¹, Xin Li², Yang Deng¹, Lidong Bing² and Wai Lam¹

¹The Chinese University of Hong Kong

²DAMO Academy, Alibaba Group

{wxzhang, ydeng, wlam}@se.cuhk.edu.hk

{xinting.lx, l.bing}@alibaba-inc.com

Abstract

Aspect-based sentiment analysis (ABSA) has received increasing attention recently. Most existing work tackles ABSA in a discriminative manner, designing various task-specific classification networks for the prediction. Despite their effectiveness, these methods ignore the rich label semantics in ABSA problems and require extensive task-specific designs. In this paper, we propose to tackle various ABSA tasks in a unified generative framework. Two types of paradigms, namely annotation-style and extraction-style modeling, are designed to enable the training process by formulating each ABSA task as a text generation problem. We conduct experiments on four ABSA tasks across multiple benchmark datasets where our proposed generative approach achieves new state-of-the-art results in almost all cases. This also validates the strong generality of the proposed framework which can be easily adapted to arbitrary ABSA task without additional task-specific model design.¹

1 Introduction

Aspect-based sentiment analysis (ABSA), aiming at mining fine-grained opinion information towards specific aspects, has attracted increasing attention in recent years (Liu, 2012). Multiple fundamental sentiment elements are involved in ABSA, including the aspect term, opinion term, aspect category, and sentiment polarity. Given a simple example sentence “*The pizza is delicious.*”, the corresponding elements are “*pizza*”, “*delicious*”, “*food quality*” and “*positive*”, respectively.

The main research line of ABSA focuses on the identification of those sentiment elements such as extracting the aspect term (Liu et al., 2015; Yin et al., 2016; Li et al., 2018; Ma et al., 2019) or classifying the sentiment polarity for a given aspect (Wang et al., 2016; Chen et al., 2017; Jiang et al., 2019; Zhang and Qian, 2020). To provide more detailed information, many recent studies propose to jointly predict multiple elements simultaneously (Li et al., 2019a; Wan et al., 2020; Peng et al., 2020; Zhao et al., 2020). Taking the Unified ABSA (UABSA, also called End-to-End ABSA) task as an example, it tries to simultaneously predict the mentioned aspect terms and the corresponding sentiment polarities (Luo et al., 2019; He et al., 2019).

In general, most ABSA tasks are formulated as either sequence-level or token-level classification problems (Li et al., 2019b). By designing task-specific classification networks, the prediction is made in a discriminative manner, using the class index as labels for training (Huang and Carley, 2018; Wan et al., 2020). However, these methods ignore the label semantics, *i.e.*, the meaning of the natural language labels, during the training process. Intuitively, knowing the meaning of “*food quality*” and “*restaurant ambiance*”, it can be much easier to identify that the former one is more likely to be the correct aspect category for the concerned aspect “*pizza*”. Such semantics of the label can be more helpful for the joint extraction of multiple sentiment elements, due to the complicated interactions of those involved elements. For example, understanding “*delicious*” is an adjective for describing the food such as “*pizza*” could better lead to the prediction of aspect opinion pair (“*pizza*”, “*delicious*”). Another issue is that different classification models are proposed to suit the need of different ABSA problems, making it difficult to adapt the model from one to another.

Motivated by recent success in formulating sev-

* Work done when Wenxuan Zhang was an intern at Alibaba. The work described in this paper is partially supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project Code: 14200719).

¹The data and code can be found at <https://github.com/IsakZhang/Generative-ABSA>

eral language understanding problems such as named entity recognition, question answering, and text classification as generation tasks (Raffel et al., 2020; Athiwaratkun et al., 2020), we propose to tackle various ABSA problems in a unified generative approach in this paper. It can fully utilize the rich label semantics by encoding the natural language label into the target output. Moreover, this unified generative model can be seamlessly adapted to multiple tasks without introducing additional task-specific model designs.

In order to enable the **Generative Aspect-based Sentiment analysis (GAS)**, we tailor-make two paradigms, namely annotation-style and extraction-style modeling to transform the original task as a generation problem. Given a sentence, the former one adds annotations on it to include the label information when constructing the target sentence; while the latter directly adopts the desired natural language label of the input sentence as the target. The original sentence and the target sentence produced by either paradigm can then be paired as a training instance of the generation model. Furthermore, we propose a prediction normalization strategy to handle the issue that the generated sentiment element falls out of its corresponding label vocabulary set. We investigate four ABSA tasks including Aspect Opinion Pair Extraction (AOPE), Unified ABSA (UABSA), Aspect Sentiment Triplet Extraction (ASTE), and Target Aspect Sentiment Detection (TASD) with the proposed unified GAS framework to verify its effectiveness and generality.

Our main contributions are 1) We tackle various ABSA tasks in a novel generative manner; 2) We propose two paradigms to formulate each task as a generation problem and a prediction normalization strategy to refine the generated outputs; 3) We conduct experiments on multiple benchmark datasets across four ABSA tasks and our approach surpasses previous state-of-the-art in almost all cases. Specifically, we obtain 7.6 and 3.7 averaged gains on the challenging ASTE and TASD task respectively.

2 Generative ABSA (GAS)

2.1 ABSA with Generative Paradigm

In this section, we describe the investigated ABSA tasks and the proposed two paradigms, namely, annotation-style and extraction-style modeling.

Aspect Opinion Pair Extraction (AOPE) aims to extract aspect terms and their corresponding

opinion terms as pairs (Zhao et al., 2020; Chen et al., 2020). Here is an illustrative example of our generative formulations for the AOPE task:

Input: Salads were fantastic, our server was also very helpful.

Target (Annotation-style): [Salads | fantastic] were fantastic here, our [server | helpful] was also very helpful.

Target (Extraction-style): (Salads, fantastic); (server, helpful)

In the annotation-style paradigm, to indicate the pair relations between the aspect and opinion terms, we append the associated opinion modifier to each aspect term in the form of [*aspect* | *opinion*] for constructing the target sentence, as shown in the above example. The prediction of the coupled aspect and opinion term is thus achieved by including them in the same bracket. For the extraction-style paradigm, we treat the desired pairs as the target, which resembles direct extraction of the expected sentiment elements but in a generative manner.

Unified ABSA (UABSA) is the task of extracting aspect terms and predicting their sentiment polarities at the same time (Li et al., 2019a; Chen and Qian, 2020). We also formulate it as an (*aspect*, *sentiment polarity*) pair extraction problem. For the same example given above, we aim to extract two pairs: (*Salads*, *positive*) and (*server*, *positive*). Similarly, we replace each aspect term as [*aspect* | *sentiment polarity*] under the annotation-style formulation and treat the desired pairs as the target output in the extraction-style paradigm to reformulate the UABSA task as a text generation problem.

Aspect Sentiment Triplet Extraction (ASTE) aims to discover more complicated (*aspect*, *opinion*, *sentiment polarity*) triplets (Peng et al., 2020):

Input: The Unibody construction is solid, sleek and beautiful.

Target (Annotation-style): The [Unibody construction | positive | solid, sleek, beautiful] is solid, sleek and beautiful.

Target (Extraction-style): (Unibody construction, solid, positive); (Unibody construction, sleek, positive); (Unibody construction, beautiful, positive);

As shown above, we annotate each aspect term with its corresponding sentiment triplet wrapped in the bracket, i.e., [*aspect*|*opinion*|*sentiment polarity*] for the annotation-style modeling. Note that

we will include all the opinion modifiers of the same aspect term within the same bracket to predict the sentiment polarities more accurately. For the extraction-style paradigm, we just concatenate all triplets as the target output.

Target Aspect Sentiment Detection (TASD) is the task to detect all (*aspect term*, *aspect category*, *sentiment polarity*) triplets for a given sentence (Wan et al., 2020), where the aspect category belongs to a pre-defined category set. For example,

Input: A big disappointment, all around. The pizza was cold and the cheese wasn’t even fully melted.

Target (Annotation-style): A big disappointment, all around. The [pizza | food quality | negative] was cold and the [cheese | food quality | negative] wasn’t even fully melted [null | restaurant general | negative].

Target (Extraction-style): (pizza, food quality, negative); (cheese, food quality, negative); (null, restaurant general, negative);

Similarly, we pack each aspect term, the aspect category it belongs to, and its sentiment polarity into a bracket to build the target sentence for the annotation-style method. Note that we use a bigram expression for the aspect category instead of the original uppercase form “FOOD#QUALITY” to make the annotated target sentence more natural. As presented in the example, some triplets may not have explicitly-mentioned aspect terms, we thus use “null” to represent it and put such triplets at the end of the target output. For the extraction-style paradigm, we concatenate all the desired triplets, including those with implicit aspect terms, as the target sentence for sequence-to-sequence learning.

2.2 Generation Model

Given the input sentence x , we generate a target sequence y' , which is either based on the annotation-style or extraction-style paradigm as described in the last section, with a text generation model $f(\cdot)$. Then the desired sentiment pairs or triplets s can be decoded from the generated sequence y' . Specifically, for the annotation-style modeling, we extract the contents included in the bracket “[]” from y' , and separate different sentiment elements with the vertical bar “[]”. If such decoding fails, *e.g.*, we cannot find any bracket in the output sentence or the number of vertical bars is not as expected,

	L14	R14	R15	R16
HAST+TOWE [†]	53.41	62.39	58.12	63.84
JERE-MHS [†]	52.34	66.02	59.64	67.65
SpanMlt (Zhao et al., 2020)	68.66	75.60	64.68	71.78
SDRN (Chen et al., 2020)	66.18	73.30	65.75	73.67
GAS-ANNOTATION-R	68.74	72.66	65.03	73.75
GAS-EXTRACTION-R	67.58	73.22	65.83	74.12
GAS-ANNOTATION	69.55	<u>75.15</u>	67.93	75.42
GAS-EXTRACTION	68.08	74.12	<u>67.19</u>	<u>74.54</u>

Table 1: Main results of the AOPE task. The best results are in bold, second best results are underlined. Results are the average F1 scores over 5 runs. [†] denotes results are from Zhao et al. (2020).

	L14	R14	R15	R16
BERT+GRU (Li et al., 2019b)	61.12	73.17	59.60	70.21
SPAN-BERT (Hu et al., 2019)	61.25	73.68	62.29	-
IMN-BERT (He et al., 2019)	61.73	70.72	60.22	-
RACL (Chen and Qian, 2020)	63.40	75.42	<u>66.05</u>	-
Dual-MRC (Mao et al., 2021)	65.94	75.95	65.08	-
GAS-ANNOTATION-R	67.37	75.77	65.75	71.87
GAS-EXTRACTION-R	66.71	76.30	64.00	72.39
GAS-ANNOTATION	68.64	<u>76.58</u>	66.78	<u>73.21</u>
GAS-EXTRACTION	<u>68.06</u>	77.13	65.96	73.64

Table 2: Main results of the UABSA task. The best results are in bold, second best results are underlined. Results are the average F1 scores over 5 runs.

we ignore such predictions. For the extraction-style paradigm, we separate the generated pairs or triplets from the sequence y' and ignore those invalid generations in a similar way.

We adopt the pre-trained T5 model (Raffel et al., 2020) as the generation model $f(\cdot)$, which closely follows the encoder-decoder architecture of the original Transformer (Vaswani et al., 2017). Therefore, by formulating these ABSA tasks as a text generation problem, we can tackle them in a unified sequence-to-sequence framework without task-specific model design.

2.3 Prediction Normalization

Ideally, the generated element $e \in s$ after decoding is supposed to exactly belong to the vocabulary set it is meant to be. For example, the predicted aspect term should explicitly appear in the input sentence. However, this might not always hold since each element is generated from the vocabulary set containing all tokens instead of its specific vocabulary set. Thus, the predictions of a generation model may exhibit morphology shift from the ground-truths, *e.g.*, from *single* to *plural* nouns.

	L14	R14	R15	R16
CMLA+ (Wang et al., 2017)	33.16	42.79	37.01	41.72
Li-unified-R (Li et al., 2019a)	42.34	51.00	47.82	44.31
Pipeline (Peng et al., 2020)	42.87	51.46	52.32	54.21
Jet (Xu et al., 2020)	43.34	58.14	52.50	63.21
Jet+BERT (Xu et al., 2020)	51.04	62.40	57.53	63.83
GAS-ANNOTATION-R	52.80	67.35	56.95	67.43
GAS-EXTRACTION-R	<u>58.19</u>	<u>70.52</u>	60.23	<u>69.05</u>
GAS-ANNOTATION	54.31	69.30	<u>61.02</u>	68.65
GAS-EXTRACTION	60.78	72.16	62.10	70.10

Table 3: Main results of the ASTE task. The best results are in bold, second best results are underlined. Results are the average F1 scores over 5 runs.

We propose a prediction normalization strategy to refine the incorrect predictions resulting from such issue. For each sentiment type c denoting the type of the element e such as the aspect term or sentiment polarity, we first construct its corresponding vocabulary set V_c . For aspect term and opinion term, V_c contains all words in the current input sentence x ; for aspect category, V_c is a collection of all categories in the dataset; for sentiment polarity, V_c contains all possible polarities. Then for a predicted element e of the sentiment type c , if it does not belong to the corresponding vocabulary set V_c , we use $\bar{e} \in V_c$, which has the smallest Levenshtein distance (Levenshtein, 1966) with e , to replace e .

3 Experiments

3.1 Experimental Setup

Datasets We evaluate the proposed GAS framework on four popular benchmark datasets including Laptop14, Rest14, Rest15, and Rest16, originally provided by the SemEval shared challenges (Pontiki et al., 2014, 2015, 2016). For each ABSA task, we use the public datasets derived from them with more sentiment annotations. Specifically, we adopt the dataset provided by Fan et al. (2019), Li et al. (2019a), Xu et al. (2020), Wan et al. (2020) for the AOPE, UABSA, ASTE, T ASD task respectively. For a fair comparison, we use the same data split as previous works.

Evaluation Metrics We adopt F1 scores as the main evaluation metrics for all tasks. A prediction is correct if and only if all its predicted sentiment elements in the pair or triplet are correct.

Experiment Details We adopt the T5 base model from *huggingface* Transformer library² for

²<https://github.com/huggingface/transformers>

	Rest15	Rest16
Baseline (Brun and Nikoulina, 2018)	-	38.10
TAS-LPM-CRF (Wan et al., 2020)	54.76	64.66
TAS-SW-CRF (Wan et al., 2020)	57.51	65.89
TAS-SW-TO (Wan et al., 2020)	58.09	65.44
GAS-ANNOTATION-R	59.27	66.54
GAS-EXTRACTION-R	<u>60.63</u>	<u>68.31</u>
GAS-ANNOTATION	60.06	67.70
GAS-EXTRACTION	61.47	69.42

Table 4: Main results of the T ASD task. The best results are in bold, second best results are underlined. Results are the average F1 scores over 5 runs.

all experiments. T5 closely follows the original encoder-decoder architecture of the Transformer model, with some slight differences such as different position embedding schemes. Therefore, the encoder and decoder of it have similar parameter size as the BERT-BASE model. For all tasks, we use similar experimental settings for simplicity: we train the model with the batch size of 16 and accumulate gradients every two batches. The learning rate is set to be $3e-4$. The model is trained up to 20 epochs for the AOPE, UABSA, and ASTE task and 30 epochs for the T ASD task.

3.2 Main Results

The main results for the AOPE, UABSA, ASTE, T ASD task are reported in Tables 1, 2, 3, 4 respectively. For our proposed GAS framework, we also present the raw results without the proposed prediction normalization strategy (with the suffix “-R”). All results are the average F1 scores across 5 runs with different random seeds.

It is noticeable that our proposed methods, based on either annotation-style or extraction-style modeling, establish new state-of-the-art results in almost all cases. The only exception is on the Rest15 dataset for the AOPE task, our method is still on par with the previous best performance. It shows that tackling various ABSA tasks with the proposed unified generative method is an effective solution. Moreover, we can see that our method performs especially well on the ASTE and T ASD tasks, the proposed extraction-style method outperforms the previous best models by 7.6 and 3.7 average F1 scores (across different datasets) on them respectively. It implies that incorporating the label semantics and appropriately modeling the interactions among those sentiment elements are essential for tackling complex ABSA problems.

	BEFORE	AFTER	LABEL
#1	<i>Bbq rib</i>	<i>BBQ rib</i>	<i>BBQ rib</i>
#2	<i>repeat</i>	<i>repeats</i>	<i>repeats</i>
#3	<i>chicken peas</i>	<i>chick peas</i>	<i>chick peas</i>
#4	<i>bodys</i>	<i>bodies</i>	None
#5	<i>cafe</i>	<i>coffee</i>	<i>coffee</i>
#6	<i>vegetarian</i>	<i>vegan</i>	<i>vegetarian</i>
#7	<i>salmon</i>	<i>not</i>	<i>spinach</i>
#8	<i>flight cookie</i>	<i>might cookie</i>	<i>fortune cookie</i>

Table 5: Example cases of the predictions before and after the prediction normalization.

3.3 Discussions

Annotation-style & Extraction-style As shown in result tables, the annotation-style method generally performs better than the extraction-style method on the AOE and UASA task. However, the former one becomes inferior to the latter on the more complex ASTE and TASP tasks. One possible reason is that, on the ASTE and TASP tasks, the annotation-style method introduces too much content, such as the aspect category and sentiment polarity, into the target sentence, which increases the difficulty of sequence-to-sequence learning.

Why Prediction Normalization Works To better understand the effectiveness of the proposed prediction normalization strategy, we randomly sample some instances from the ASTE task that have different raw prediction and normalized prediction (*i.e.*, corrected by our strategy). The predicted sentiment elements before and after the normalization, as well as the gold label of some example cases are shown in Table 5. We find that the normalization mainly helps on two occasions: The first one is the morphology shift where two words have minor lexical differences. For example, the method fixes “*Bbq rib*” to “*BBQ rib*” (#1) and “*repeat*” to “*repeats*” (#2). Another case is orthographic alternatives where the model might generate words with the same etyma but different word types, *e.g.*, it outputs “*vegetarian*” rather than “*vegan*” (#6). Our proposed prediction normalization, which finds the replacement from the corresponding vocabulary set via Levenshtein distance, is a simple yet effective strategy to alleviate this issue.

We also observe that our prediction strategy may fail if the raw predictions are quite lexically different or even semantically different from the gold-standard labels (see Case #4, #7 and #8). In these

cases, the difficulty does not come from the way of performing prediction normalization but the generation of labels close to the ground truths, especially for the examples containing implicit aspects or opinions (Case #4).

4 Conclusions and Future Work

We tackle various ABSA tasks in a novel generative framework in this paper. By formulating the target sentences with our proposed annotation-style and extraction-style paradigms, we solve multiple sentiment pair or triplet extraction tasks with a unified generation model. Extensive experiments on multiple benchmarks across four ABSA tasks show the effectiveness of our proposed method.

Our work is an initial attempt on transforming ABSA tasks, which are typically treated as classification problems, into text generation problems. Experimental results indicate that such transformation is an effective solution to tackle various ABSA tasks. Following this direction, designing more effective generation paradigms and extending such ideas to other tasks can be interesting research problems for future work.

References

- Ben Athiwaratkun, Cícero 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 2020*, pages 375–385.
- Caroline Brun and Vassilina Nikoulina. 2018. [Aspect based sentiment analysis into the wild](#). In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2018*, pages 116–122.
- Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. [Recurrent attention network on memory for aspect sentiment analysis](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017*, pages 452–461.
- Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. 2020. [Synchronous double-channel recurrent network for aspect-opinion pair extraction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020*, pages 6515–6524.
- Zhuang Chen and Tiejun Qian. 2020. [Relation-aware collaborative learning for unified aspect-based sentiment analysis](#). In *Proceedings of the 58th Annual*

- Meeting of the Association for Computational Linguistics, ACL 2020, pages 3685–3694.
- Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. [Target-oriented opinion words extraction with target-fused neural sequence labeling](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, pages 2509–2518.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. [An interactive multi-task learning network for end-to-end aspect-based sentiment analysis](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019*, pages 504–515.
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. [Open-domain targeted sentiment analysis via span-based extraction and classification](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019*, pages 537–546.
- Binxuan Huang and Kathleen M. Carley. 2018. [Parameterized convolutional neural networks for aspect level sentiment classification](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1091–1096.
- 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 2019*, pages 6279–6284.
- Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals.
- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019a. [A unified model for opinion target extraction and target sentiment prediction](#). In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019*, pages 6714–6721.
- Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. [Aspect term extraction with history attention and selective transformation](#). In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 4194–4200.
- Xin Li, Lidong Bing, Wenxuan Zhang, and Wai Lam. 2019b. [Exploiting BERT for end-to-end aspect-based sentiment analysis](#). In *Proceedings of the 5th Workshop on Noisy User-generated Text, W-NUT@EMNLP 2019*, pages 34–41.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies.
- Pengfei Liu, Shafiq R. Joty, and Helen M. Meng. 2015. [Fine-grained opinion mining with recurrent neural networks and word embeddings](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015*, pages 1433–1443.
- Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang. 2019. [DOER: dual cross-shared RNN for aspect term-polarity co-extraction](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019*, pages 591–601.
- Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. [Exploring sequence-to-sequence learning in aspect term extraction](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019*, pages 3538–3547.
- Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. [A joint training dual-mrc framework for aspect based sentiment analysis](#). *CoRR*, abs/2101.00816.
- Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. [Knowing what, how and why: A near complete solution for aspect-based sentiment analysis](#). In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 8600–8607.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia V. Loukachevitch, Evgeniy V. Kotelnikov, Núria Bel, Salud María Jiménez Zafra, and Gülsen Eryigit. 2016. [Semeval-2016 task 5: Aspect based sentiment analysis](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016*, pages 19–30.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. [Semeval-2015 task 12: Aspect based sentiment analysis](#). In *SemEval@NAACL-HLT*, pages 486–495.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. [Semeval-2014 task 4: Aspect based sentiment analysis](#). In *SemEval@COLING 2014*, pages 27–35.
- 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](#). *J. Mach. Learn. Res.*, 21:140:1–140:67.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural*

Information Processing Systems 2017, pages 5998–6008.

Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z. Pan. 2020. [Target-aspect-sentiment joint detection for aspect-based sentiment analysis](#). In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 9122–9129.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. [Coupled multi-layer attentions for co-extraction of aspect and opinion terms](#). In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 3316–3322.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. [Attention-based LSTM for aspect-level sentiment classification](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016*, pages 606–615.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. [Position-aware tagging for aspect sentiment triplet extraction](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020*, pages 2339–2349.

Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. [Unsupervised word and dependency path embeddings for aspect term extraction](#). In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016*, pages 2979–2985.

Mi Zhang and Tieyun Qian. 2020. [Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020*, pages 3540–3549.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. [SpanMlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020*, pages 3239–3248.