

Exploring Conditional Text Generation for Aspect-Based Sentiment Analysis

Siva Uday Sampreeth Chebolu¹, Franck Deroncourt², Nedim Lipka², Thamar Solorio¹

¹University of Houston, USA

²Adobe Research

¹{sivauday.sampreeth8, thamar.solorio}@gmail.com

²{franck.deroncourt@gmail.com, lipka@adobe.com}

Abstract

Aspect-based sentiment analysis (ABSA) is an NLP task that entails processing user generated reviews to determine (i) the target being evaluated, (ii) the aspect category to which it belongs, and (iii) the sentiment expressed towards the target and aspect pair. In this article, we propose transforming ABSA into an abstract summary-like conditional text generation task that uses targets, aspects, and polarities to generate auxiliary statements. To demonstrate the efficacy of our task formulation and a proposed system, we fine-tune a pre-trained model for conditional text generation task to get new state-of-the-art results on a few restaurant domain and urban neighborhoods domain benchmark datasets.

1 Introduction

Consumers and product makers/service providers both benefit from user-generated evaluations on e-commerce platforms. Reading about previous customer experiences can assist future customers in making decisions. Reading about the aspects that create user feedback may help manufacturers and merchants develop ways to increase customer happiness. Thousands of reviews covering various aspects and their corresponding opinions may be found in many situations. Aspects can be a feature, a characteristic, or a behavior of a product or an entity, such as the ambiance of a restaurant, the performance of a laptop, the display of a phone, and so on. Aspect-based sentiment analysis is a task in which the sentiment for each aspect of an entity is determined.

This problem has two sub-problems: 1) aspect extraction (for example, sushi, pasta, and well-behaved staffs) and 2) finding the polarity toward each aspect. Aspect extraction involves two sub-tasks: a) extracting aspect terms and b) categorizing/normalizing the extracted aspect terms into aspect categories. Additionally, there are two sub-tasks in polarity detection: a) identify the polarity of an aspect word, b) determine the polarity of each category and build triplets from sentences [aspect term, aspect category, polarity].

To identify the aspect categories and their sentiments, Sun et al. (2019) recently turned this task into a sentence-pair classification task, such as Question-Answering or Natural Language Inference. Wan et al. (2020), on the other hand, pointed out that sentiment is affected by both the aspect category and the aspect terms or targets in a review. To overcome this challenge, they also proposed a joint model to extract the targets, aspect categories, and polarity by fine-tuning the BERT model with the auxiliary phrases (aspect-sentiment) generated by Sun et al. (2019), as well as a sequence classification to identify the targets.

The drawback of such a formulation is that it is unable to detect implicit targets in situations where the same aspect-category-sentiment pair has explicit targets associated with it. Take this review for example: *"Always crowded, but they are good at seating you promptly and have quick service."* [{service, SERVICE#GENERAL, positive}, {NULL, SERVICE#GENERAL, positive}] are the actual labels for the targets, categories, and polarities. Sun et al. (2019), on the other hand, deleted the second opinion owing to redundancy, while Wan et al. (2020)

removed it due to target conflicts.

Recently, Google released a unified framework (T5) Raffel et al. (2020) that transforms all text-based language problems into a text-to-text format, achieving state-of-the-art results on a variety of benchmarks including summarization, question answering, text classification, and more. We suggest utilizing the multi-task method offered by Wan et al. (2020) to reap the benefits of the T5 architecture. Deep generative models offer more expressive power, the ability to handle more forms of data, and the inherent benefit of being able to produce samples from the input text, all of which point to their suitability for the task at hand. The major benefit of conditional text generation is that it is mostly dependent on the content of an input text, which in our case contributes to the benefit of knowing which aspect category or polarity label to give and, finally, identify the target of those categories.

In this work, we look at a few different ways to create an auxiliary phrase and turn ABSA into a conditional text generation task based on aspects and polarities, akin to abstractive summarization. On the ABSA challenge, we fine-tune the pre-trained models from T5, and BART (Lewis et al. (2019)) and reach new state-of-the-art results. We also perform comparison tests to demonstrate that generation based on the multi-task framework is superior than specific task fine-tuned individually, implying that the improvement is due to both the pre-trained model and our approach. Furthermore, we investigate the performance of the T5-Encoder, which is pre-trained on a larger dataset than BERT, in the architecture proposed by Wan et al. (2020), to show that the improvement is due to the conversion to text generation rather than the larger dataset it is trained on. We observed that, this formulation enhanced implicit target identification while also removing the target conflict constraint in the sentence-pair classification approach proposed by Wan et al. (2020).

The following are our major contributions:

- We construct the ABSA task as a conditional text generation problem, which comprises jointly extracting targets, categorizing the aspects into pre-defined categories, and their polarity.
- All of ABSA’s sub-problems and associated

tasks may be addressed in a sequence-to-sequence framework with our re-formulation, utilizing pre-trained models like T5.

- We do comprehensive tests on four public datasets, each of which comprises a subset of the ABSA subtasks, and show that our proposed framework outperforms current state-of-the-art techniques by a considerable margin.

2 Methodology

2.1 Task Description

The T ASD task aims to detect all triples (t, a, p) that S entails in the natural language meaning, where t (called a target) is a subsequence of S , a is an aspect in A , and p is a sentiment polarity (simply called a sentiment) in P , given a sentence S consisting of n words s_1, \dots, s_n , a predefined set A of aspects and a predefined set P of polarities or sentiments. The target t can be *NULL*, indicating that it is empty. An implicit target case is the name given to this situation. We name the triple (t, a, p) an opinion since the overall objective of ABSA is to identify fine-grained sentiments. Target-Detection (TD), Aspect-Detection (AD), Target-Sentiment Joint Detection (TSD), Aspect-Sentiment Joint Detection (ASD), and Target-Aspect Joint Detection (TAD) are the natural subtasks that arise from the T ASD task. From the example in Section 1, there are 2 opinions ($\{\text{service, SERVICE\#GENERAL, positive}\}$) and ($\{\text{NULL, SERVICE\#GENERAL, positive}\}$). With two positive sentiments, there is just one aspect for the two opinions, *SERVICE\#GENERAL*, one explicit target, *service*, and one implicit target, *NULL*.

Apart from the tasks listed above, ABSA has two simpler tasks: one aims to classify sentiment according to a given aspect, as investigated in Wang et al. (2016); Sun et al. (2019), and the other aims to classify sentiment according to a given target, as studied in Zeng et al. (2019). These tasks are not comparable to the T ASD or its sub-tasks because they rely on prerequisite tasks like AD or TD to complete ABSA.

2.2 Construction of auxiliary sentence

To transform the ABSA task into a conditional text generation task, we explored the following two methods:

Task	Dataset	Pseudo Phrase	Pseudo Sentence
AD	All	aspect	The review expressed opinion on [aspect]
ASD	All	aspect~polarity	The review expressed [polarity] opinion on [aspect]
TD	SE-15, SE-16	target	The review expressed opinion for [target]
TSD	SE-15, SE-16	target~polarity	The review expressed [polarity] opinion for [target]
TAD	SE-15, SE-16	target~aspect	The review expressed opinion on [aspect] for [target]
TASD	SE-15, SE-16	target~aspect~polarity	The review expressed [polarity] opinion on [aspect] for [target]

Table 1: Auxiliary Output Sentence formats and datasets on which the tasks are applied

Pseudo output phrase: The sentence we obtain from a text generation model will only have the *targets*, *aspects*, and *polarities* separated by a delimiter. For the above example, we expect a model to generate (*service* ~ *SERVICE#GENERAL* ~ *positive* ~ *NULL* ~ *SERVICE#GENERAL* ~ *positive*).

Pseudo output sentence: We will generate a complete sentence using a pseudo sentence format to obtain the labels for aspects, targets, and polarities. The pseudo sentence for the same example will be “*The review expressed [positive] opinion on [SERVICE#GENERAL] for [service], [positive] opinion on [SERVICE#GENERAL] for [NULL]*”.

The modifications in the format of the pseudo output phrase/sentence for all TASD subtasks are shown in Table 1. Once we’ve created the auxiliary phrase/sentence for the output of a generation model, we can transform the ABSA task from a sentence classification task to a conditional text generation task. This is a necessary step that can significantly improve the ABSA task’s experimental results.

2.3 Fine-tuning pre-trained T5 model

T5 is a huge new neural network model that is trained on a combination of unlabeled text (the new C4 collection of English online text) and labeled data from popular natural language processing tasks, then fine-tuned separately for each task. It’s an encoder-decoder model that translates all NLP problems to text-to-text. It requires an input sequence and a target sequence for training and is trained via teacher forcing. To do this, we feed the review text into the model as an input sequence and train it to generate a target sequence with one of the auxiliary sentence styles. To do this, we feed the review text into the model as an input sequence and train it to generate a target sequence with one of the auxiliary sentence styles. During the prediction phase, we simply pro-

vide the review text and evaluate the model output by extracting the targets, aspect categories, and polarity.

2.4 T5-Joint and T5-Separate

T5-Joint: This is fine-tuning the T5 model for concurrently detecting targets, aspect categories, and polarities in order to complete all six tasks. As a fine-tuning target sequence, we utilize the format in the final row of Table 1.

T5-Separate: Here, we fine-tune the T5 model to solve the six tasks independently with a separate auxiliary sentence/phrase formats, as mentioned in Table 1, for each task.

3 Experiments

3.1 Datasets

The datasets that we utilized to evaluate our model are described in this section.

SE-14: SemEval-2014 Task4 dataset on restaurant reviews is used to evaluate the AD and ASD tasks. Our model solves sub-task 3 (Aspect Category Detection) and the joint detection of aspect-sentiment tasks. It is note that the sentiment score reported does not correspond to the subtask-4 (Aspect Category Sentiment Analysis) on the SE-14 dataset. This is because the formulation of the subtask-4 assumes that the gold aspect categories are already available for sentiment prediction, which is not the case for the task at hand.

SE-15 and SE-16: We also did tests on two other datasets in the restaurant domain, one of which is from SemEval-2015 (SE-15) Task 12, and the other is from SemEval-2016 (SE-16) Task 5. On these two datasets, we assess all six tasks. We also assess TSD and TASD tasks for implicit targets, as described in Wan et al. (2020).

SH: We evaluate on the Sentihood (SH) dataset, which describes locations or neighborhoods of Lon-

don and was collected from question answering platform of Yahoo. The sentiment polarities are $p \in P = \{\text{positive, negative and none}\}$, the targets are $t \in T = \{\text{Location1, and Location2}\}$, and the aspect categories are $a \in A = \{\text{general, price, transit-location, and safety}\}$. The definition of target in this dataset, however, differs from that of the other benchmarks. The aspect-category is formed by combining the target and the aspect, resulting in eight aspect categories from the provided targets and aspects. As a result, we only evaluate the AD and ASD tasks on this dataset, rather than all six.

3.2 Hyperparameters and Metrics

For all of our T5 model fine-tuning studies, we use the pre-trained T5-Base model. The encoder and decoder each include 12 transformer blocks, the size of the hidden layer is 768, and the pre-trained model has 220 million parameters. We utilize a maximum input sequence length of 128, a train batch size of 16, an evaluation batch size of 64, and a learning rate of $4e-5$ for fine-tuning with a maximum of 50 epochs. Micro-F1 score and accuracy are used to evaluate our results for SE-14, SE-15, SE-16, and MAMS-ACSA, whereas macro-F1 score is used for the SH dataset.

3.3 Comparison Methods

We compare our model with the following models:

SemEval-Top The best scores in the SemEval contests are represented by SemEval-Top. These scores are available for three subtasks: AD, TD, and TAD.

MTNA MTNA Xue et al. (2017) is a multi-task model that identifies both aspect categories and targets using RNN and CNN. For both AD and TD, the research reported results on SE-14, SE-15, and SE-16.

Sentic LSTM + TA + SA Ma et al. (2018) investigated Hierarchical attention, which first attends to the targets in a given review and then combines the aspect-embeddings and the output of a Sentic-LSTM network to apply another level of attention in order to produce target-aspect-specific sentence representations. They reported results on SE-15 and SH datasets on AD and ASD tasks.

TAN Instead of using aspect categories, Movahedi et al. (2019) created topic-specific sentence representations based on numerous topics. Furthermore, they employ a regularization term similar to Hu et al. to

assure the uniqueness of the topics. They reported results on SE-14 and SE-16 for AD task only.

BERT-pair-NLI-B To fine-tune the pre-trained model using BERT for AD and ASD tasks, Sun et al. (2019) generate an auxiliary sentence from the aspect and transform ABSA to a sentence-pair classification task. They report results on SE-14 and SH datasets.

baseline-1-f.lex Brun and Nikoulina (2018) used a T ASD pipeline technique. The study only presented findings on SE-15 for both T ASD and its subtask ASD, and no source code is supplied.

DE-CNN Xu et al. (2018) used a CNN model for TD. The paper only reported results on SE-16.

THA + STN A neural model for TD that includes a bi-linear attention layer and an FC layer Li et al. (2018b). Both SE-15 and SE-16 were reported in the study.

BERT-PT A BERT-based model to extract the targets (TD) Xu et al. (2019). Results are taken from Wan et al. (2020) for SE-15 and SE-16.

E2E-TBSA Two layered recurrent neural networks are used in a unified framework to solve TSD in an end-to-end manner Li et al. (2019): The upper one predicts the unified tags, while the bottom one predicts the auxiliary target boundaries. Results are obtained from Wan et al. (2020) for SE-15 and SE-16.

DOER A dual cross-shared RNN model for TSD Luo et al. (2019). We used the results published in Wan et al. (2020) for SE-15 and SE-16.

TAS-BERT-SW-TO Wan et al. (2020) improved the BERT-pair-NLI-B architecture by detecting the presence of targets and extracting them using a TO tag-schema along with the aspect and sentiment of a review sentence. They reported results on SE-15 and SE-16 datasets for all the six tasks.

TAS-BERT-SW-BIO-CRF Wan et al. (2020) improved the BERT-pair-NLI-B architecture by detecting the presence of targets and extracting them using a BIO tag-schema along with the aspect and sentiment of a review sentence. They reported results on SE-15 and SE-16 datasets for all the six tasks.

QACG-BERT Wu and Ong (2020) create a CG-BERT that employs context-guided (CG) softmax-attention by first adapting a context-aware Transformer. Next, they present an enhanced Quasi-Attention CG-BERT model that learns a compositional attention model that enables subtractive attention.

4 Results

Model	AD (Micro F1)	ASD (Accuracy)		
		4-way	3-way	2-way
SemEval-Top	88.58	82.90	-	-
MTNA	88.91	-	-	-
TAN	90.61	-	-	-
SCAN-BERT-AVE	-	88.61	-	-
BERT-pair-NLI-B	92.18	78.65	79.98	84.35
QACG-BERT	92.64	77.80	80.10	82.77
T5-Phrase-AD	85.30	-	-	-
T5-Sentence-AD	92.50	-	-	-
T5-Phrase-Joint-ASD	91.00	79.00	81.00	82.00
T5-Sentence-Joint-ASD	93.34	82.75	84.50	85.62

Table 2: Results for ASD task on SE-14 dataset. Micro F1-score is used for AD and Accuracy for ASD task.

Model	AD (Macro F1)	ASD (Accuracy)
LSTM-Final	68.90	82.00
Sentic LSTM + TA + SA	78.20	89.30
Dmu-Entnet	78.50	91.00
BERT-pair-QA-M	86.40	93.60
BERT-pair-QA-B	87.90	93.30
QACG-BERT	89.70	93.80
T5-Phrase-AD	89.68	-
T5-Sentence-AD	88.75	-
T5-Phrase-Joint-ASD	90.45	93.06
T5-Sentence-Joint-ASD	90.63	91.56

Table 3: Results on SentiHood dataset. Macro F1-score is used for AD and Accuracy for ASD task.

Results on SE-14, SE-15 and SE-16, and SH datasets are presented in Table 2, Table 4, and Table 3 respectively. Each experiment’s T5-based scores are averaged over three runs. Each table’s first section relates to the results of earlier studies on that dataset. The ablation study for a subtask of the main task described in that table is the second portion. For instance, ASD is the core task in the SE-14 and SH datasets, whilst the AD is utilized as an ablation experiment to assess the relevance of joint detection. Similarly, the main task in the SE-15 and SE-16 datasets is TASD, with the subtasks ASD, TSD, and TAD being examined. We report AD and ASD, and TD and TSD, for ASD and TSD respectively. Further, we report AD and TD scores for TAD. Finally, we provide the joint detection scores of each job for the specified dataset in the table’s last section. The remainder of this section will use a task-based classification to describe each of those results.

On the Restaurant and Urban Neighborhood domain datasets, our conditional text generation-based

model consistently outperforms all attention-based and transformer-based approaches. These findings show that our technique correctly interprets the input language in order to generate the appropriate aspect categories, targets, and polarities. Furthermore, the T5-based studies demonstrate that performance on the complex implicit opinions is much enhanced. Our ablation task experiment and primary task findings (TASD), which included AD and TD, beat the TAS-BERT model considerably. Furthermore, we can see that joint modeling experiments produce better results than single-task studies. Moreover, jobs that included target identification improved by at least 3%.

On both ablation and primary tasks, the pseudo-sentence based formulation beat the phrase-based formulation in the SE-14 dataset. The trend is inconsistent on the SE-15 and SE-16 datasets, though. The phrase-based formulation outperformed the sentence-based formulation on the majority of target-detection tasks. The phrase-based generation, for example, had the best performance on the TD, TSD, and TASD tasks on both datasets. On a high level, phrase-based formulation did better on the SE-15 dataset, whereas sentence-based formulation performed better on aspect-detection related tasks in SE-16. It is possible that the longer sequence that had to be generated for the sentence-based formulation with multiple opinions had an impact on performance on target-related tasks.

4.1 Qualitative Analysis

We observed different types of errors that occurred due to wrong predictions from the models. We present a few case studies of these errors in Table 5, which we discuss in this section. The first error made by the models is due to the insufficient context provided by the review sentence. For example, the sentence S3 in Table 5 is only a part of an entire review that has "Green Tea creme brulee is a must!" as the previous sentence. The predicted aspect category *RESTAURANT#GENERAL* is true when we consider the S3 independently. However, when we take the previous sentence into context, the gold annotation will be valid. Similar is the case with S3, for which the models predict *FOOD#QUALITY* as an aspect category, assuming that the word *it* refers to some food item. This drawback historically exists in the

AD			TD			TAD		
Method	SE-15	SE-16	Method	SE-15	SE-16	Method	SE-15	SE-16
SemEval-Top	62.68	73.03	SemEval-Top	70.05	72.34	SemEval-Top	42.90	52.61
MTNA	65.97	76.42	MTNA	67.73	72.95	TAS-BERT-SW-BIO-CRF	63.37	71.64
Sentic LSTM + TA + SA	73.82	-	DE-CNN	-	74.37	TAS-BERT-SW-TO	62.60	69.98
TAN	-	73.38	THA+STN	71.46	73.61	TAS-BERT-large-SW-BIO-CRF	63.97	72.45
BERT-pair-NLI-B	70.78	80.25	BERT-PT	73.15	77.97	TAS-BERT-large-SW-TO	63.31	70.94
TAS-BERT-SW-BIO-CRF	76.34	81.57	TAS-BERT-SW-BIO-CRF	75.00	81.37			
TAS-BERT-SW-TO	76.40	82.77	TAS-BERT-SW-TO	71.54	78.10			
TAS-BERT-large-SW-BIO-CRF	77.18	82.05	TAS-BERT-large-SW-BIO-CRF	76.13	81.99			
TAS-BERT-large-SW-TO	77.32	83.64	TAS-BERT-large-SW-TO	73.06	79.22			
BART-Phrase-AD	71.13	85.53	BART-Phrase-TD	78.45	79.51			
T5-Phrase-AD	72.96	74.53	T5-Phrase-TD	-	-			
T5-Sentence-AD	79.25	83.75	T5-Sentence-TD	78.46	79.53			
BART-Phrase-ASD	80.70	84.56	BART-Phrase-TSD	76.29	77.13	BART-Phrase-TAD	63.42	67.17
T5-Phrase-ASD	79.45	84.05	T5-Phrase-TSD	79.38	83.32	T5-Phrase-TAD	67.53	74.07
T5-Sentence-ASD	79.11	85.24	T5-Sentence-TSD	79.60	82.02	T5-Sentence-TAD	66.48	73.02
BART-Phrase-TAD	78.99	84.28	BART-Phrase-TAD	80.73	82.76			
T5-Phrase-TAD	77.43	82.97	T5-Phrase-TAD	80.73	82.76			
T5-Sentence-TAD	77.98	84.62	T5-Sentence-TAD	80.65	84.30			
BART-Phrase-Joint-TASD	78.85	84.19	BART-Phrase-Joint-TASD	76.99	76.06	BART-Phrase-Joint-TASD	64.63	66.49
T5-Phrase-Joint-TASD	77.31	83.10	T5-Phrase-Joint-TASD	80.69	83.51	T5-Phrase-Joint-TASD	67.19	74.65
T5-Sentence-Joint-TASD	78.58	82.97	T5-Sentence-Joint-TASD	79.98	83.54	T5-Sentence-Joint-TASD	67.72	73.03

ASD			TSD			TASD		
Method	SE-15	SE-16	Method	SE-15	SE-16	Method	SE-15	SE-16
baseline-1_f_lex	-	63.50	E2E-TBSA	53.00 (-)	63.10 (-)	baseline-1_f_lex	-	38.10
BERT-pair-NLI-B	63.67	72.70	DOER	56.33 (-)	65.91 (-)	TAS-BERT-SW-BIO-CRF	57.51	65.89
TAS-BERT-SW-BIO-CRF	68.50	74.12	TAS-BERT-SW-BIO-CRF	66.11 (64.29)	75.68 (72.92)	TAS-BERT-SW-TO	58.09	65.44
TAS-BERT-SW-TO	70.42	76.33	TAS-BERT-SW-TO	64.84 (65.02)	73.34 (71.02)	TAS-BERT-large-SW-BIO-CRF	58.12	67.19
TAS-BERT-large-SW-BIO-CRF	69.12	74.87	TAS-BERT-large-SW-BIO-CRF	67.23 (66.09)	75.96 (76.42)	TAS-BERT-large-SW-TO	58.53	66.29
TAS-BERT-large-SW-TO	68.95	75.81	TAS-BERT-large-SW-TO	63.99 (66.19)	72.74 (74.00)	TAS-T5-SW-TO	52.08	59.63
BART-Phrase-ASD	70.98	73.80	BART-Phrase-TSD	71.52 (70.80)	68.16 (68.34)	TAS-T5-SW-BIO-CRF	53.29	61.56
T5-Phrase-ASD	71.65	78.09	T5-Phrase-TSD	70.95 (71.46)	78.03 (77.40)			
T5-Sentence-ASD	71.00	79.00	T5-Sentence-TSD	71.04 (70.82)	76.75 (76.03)			
BART-Phrase-Joint-TASD	71.24	75.71	BART-Phrase-Joint-TASD	68.92 (68.91)	68.80 (68.00)	BART-Phrase-Joint-TASD	58.53	60.25
T5-Phrase-Joint-TASD	70.13	77.17	T5-Phrase-Joint-TASD	71.60 (72.16)	77.07 (78.22)	T5-Phrase-Joint-TASD	61.42	69.85
T5-Sentence-Joint-TASD	70.55	76.05	T5-Sentence-Joint-TASD	71.57 (70.82)	76.50 (77.20)	T5-Sentence-Joint-TASD	61.15	67.48

Table 4: Results for SE-15 and SE-16 datasets for six tasks.

NLP problems, and using coreference resolution is one solution to overcome this.

The second error is not correctly identifying the targets. In the sentence S1, the target *food* belongs to both the opinions *mediocre* and the *price*. TAS-BERT model wasn't able to identify the target *food*, which is farther away from the opinion on price, for the aspect category *FOOD#PRICES*. This can be attributed to the performance improvement in the target detection by the generation models. However, in sentence S2, the BART-based model wasn't able to identify the second target completely. Even though TAS-BERT could identify a second target, it was not able to recognize that they are two separate targets and should have been two different opinions. The T5-based model can identify all the targets, aspects, and sentiments correctly. This is the third error type that we observed during our analysis.

The last error we noticed is that, even though the models identified the correct target-aspect pair, the sentiment given to that pair is wrong. For example,

in sentence S3, all three methods identified the target-aspect pair correctly. However, the TAS-BERT and BART-based models made wrong sentiment predictions. We can also attribute these wrong sentiment predictions to the difference in the performance of the respective models on TAD, and TASD tasks from Table 4.

4.2 Comparison of BERT, BART, and T5

BART-based generation outperform BERT-based approaches on all tasks in the SE-15 dataset and several tasks in the SE-16 dataset, as shown in Table 4. With the exception of aspect identification on the SE-15 dataset, T5-based generation consistently beat BART-based techniques in all situations. Furthermore, we do not report the results of the BART model's pseudo-sentence-based generation because it performed badly when compared to the phrase-based formulation due to its weak generalization capacity to the format we used for the predictions.

We wanted to make sure that the T5 model's im-

Id	Review Sentence	Gold Annotations	Model & Predictions		
			TAS-BERT	BART	T5
S1	My g/f and I both agreed the food was very mediocre especially considering the price.	{FOOD#PRICES, food, negative} {FOOD#QUALITY, food, negative}	{FOOD#PRICES, NULL, negative} {FOOD#QUALITY, food, neutral}	{FOOD#PRICES, food, negative} {FOOD#QUALITY, food, negative}	{FOOD#PRICES, food, negative} {FOOD#QUALITY, food, negative}
S2	Amazing Spanish Mackerel special appetizer and perfect box sushi (that eel with avodcao – um um um).	{FOOD#QUALITY, Spanish Mackerel special appetizer, positive} {FOOD#QUALITY, box sushi, positive} {FOOD#QUALITY, eel with avodcao, positive}	{FOOD#QUALITY, Spanish Mackerel special appetizer, positive} {FOOD#QUALITY, perfect box sushi (that eel with avodcao, positive)} Missing	{FOOD#QUALITY, Spanish Mackerel special appetizer, positive} Missing Missing	{FOOD#QUALITY, Spanish Mackerel special appetizer, positive} {FOOD#QUALITY, box sushi, positive} {FOOD#QUALITY, eel with avodcao, positive}
S3	But the space is small and lovely, and the service is helpful	{AMBIENCE#GENERAL, space, positive'} {SERVICE#GENERAL, service, positive'}	{AMBIENCE#GENERAL, space, negative'} {SERVICE#GENERAL, service, positive'}	{AMBIENCE#GENERAL, space, negative'} {SERVICE#GENERAL, service, positive'}	{AMBIENCE#GENERAL, space, positive'} {SERVICE#GENERAL, service, positive'}
S4	Don't leave the restaurant without it.	{FOOD#QUALITY, NULL, positive}	{RESTAURANT#GENERAL, NULL, positive}	{RESTAURANT#GENERAL, NULL, negative}	{RESTAURANT#GENERAL, NULL, positive}
S5	It was absolutely amazing.	{RESTAURANT#GENERAL, NULL, positive}	{FOOD#QUALITY, NULL, positive}	{FOOD#QUALITY, NULL, positive}	{FOOD#QUALITY, NULL, positive}

Table 5: Qualitative Analysis. Wrong predictions are highlighted in **Red (bold)**. Correct predictions are highlighted in *Green (italicized and lighter font)*

improvements were attributable to our formulation and use of conditional text generation to address the problem, rather than the dataset’s size or the number of steps it was pre-trained on. As a result, we conducted an experiment in which the T5-Encoder was used in place of the BERT-Encoder in the TAS-BERT model. For the SE-15 and SE-16 datasets, the results are only presented for the primary TASD job. It is clear that our model outperformed the TAS-T5-Encoder-based model substantially. This might be because T5 has been pre-trained on a text-to-text formulation rather than encoder-only classification tasks. Furthermore, we used BERT-large, which has 340M parameters, to offer a fair comparison to the T5 model, which has 220M parameters vs 110M for BERT-base. Table 4 shows that BERT-large, with more than twice the amount of parameters as BERT-base, did not produce compelling improvements. With more than 1.5 times the amount of parameters, T5 surpassed the BERT-large model once more. This demonstrates that the amount of parameters has no bearing on T5’s performance.

5 Related Work

In this section, we take a look at the background of conditional text generation based pre-trained models and existing work on the different sub-tasks of ABSA.

5.1 Conditional Text Generation

Text summarization Rush et al. (2015), reading comprehension and other applications of text generation from text to text may be found. Natural language text is generated for communicating, summarizing, or refining by comprehending the source text and getting its semantic representation. In recent years, pre-trained language models have become a trend for various tasks in NLP, mainly for text generation, and conditional text generation. From the available pre-trained models for language generation Liu et al. (2019); Radford et al. (2019); Lewis et al. (2019); Raffel et al. (2020), we chose to use BART, and T5 in this work. Both BART and T5 employ a conventional Transformer design (Encoder-Decoder), similar to the original Transformer model used for neural machine translation, but with some differences. BART is more like a combination of BERT (which only uses the encoder) Devlin et al. (2019) and GPT (which only uses the decoder) Radford et al. (2019). The encoder utilizes a denoising goal similar to BERT, while the decoder tries to recreate the original sequence (autoencoder) token by token using the preceding (uncorrupted) tokens and the encoder output. T5 uses common crawl web extracted text. The pre-trained model is trained for 2^{19} steps using a BERT-base size encoder-decoder transformer with the denoising goal and the C4 dataset. T5 is pre-trained on following ob-

jectives: (1) language modeling (predicting the next word), (2) BERT-style objective (masking/replacing words with random words and predicting the original text), and (3) deshuffling (which is shuffling the input randomly and try to predict the original text).

5.2 ABSA Sub-tasks

5.2.1 Individual tasks

AD A lot of research has been done on the AD task over the years, which seeks to detect aspects only. SVM classifiers are trained to detect features in earlier research such as Kiritchenko et al. (2014); Alghunaim (2015). Liu et al. (2018) provides neural models to help with the AD task performance. Different attention processes are incorporated to a neural model by Xue et al. (2017); Movahedi et al. (2019); Rani and Subramanian (2020); Hu et al. (2019a); Liang et al. (2019) to identify aspects more accurately. Recently, Sun et al. (2019); Li et al. (2020); Wan et al. (2020) used pre-trained models such as BERT to get encoded representations of different types of input to identify the aspect categories.

TD The TD task aims to extract targets only. Traditionally, CRFs are used by Jakob and Gurevych (2010); Yin et al. (2016) to extract the targets. Liu et al. (2013) used syntactic patterns. Lately, neural models such as CNN Xue et al. (2017); Xue and Li (2018) and RNN Li and Lam (2017); Xue et al. (2017); Luo et al. (2018) are widely used in target extraction. Different attention mechanisms are introduced by Wang et al. (2017); Li et al. (2018b), in addition to a neural model to extract targets. Wan et al. (2020) used a CRF and a softmax decoding strategy on the encoded token representation from the pre-trained model BERT to detect the targets.

5.2.2 Joint tasks

ASD The ASD task is intended to identify both aspects and sentiments at the same time. Schmitt et al. (2018) uses an end-to-end CNN model to handle this problem. This task is converted into a sentence-pair classification problem using Sun et al. (2019), which allows the BERT model to be fine-tuned. Wan et al. (2020), which was released recently, takes a similar technique to solving this task in a joint modeling scenario.

TSD The TSD job tries to identify targets and sentiments at the same time. Mitchell et al. (2013)

and Zhang et al. (2015) simplify the TSD job to a sequence labeling problem, which is solved by a CRF decoder using hand-crafted linguistic features, respectively. Recently, neural models have been popular, and Li et al. (2018a) presents a unified model for the TSD problem that consists of two stacked LSTM networks. Luo et al. (2019) proposes a dual cross-shared RNN model for TSD that incorporates sentiment lexicon and part-of-speech information. Hu et al. (2019b) offers a span-based pipeline architecture for solving the TSD challenge by fine-tuning a language model that has already been trained. More recently, Wan et al. (2020) solved this problem using a joint modelling objective by fine-tuning BERT on the input representation and using a softmax/CRF decoding mechanism on the encoded token representations.

TAD The TAD task aims to detect the aspect categories and the target simultaneously. There is very limited work on this task. In this task, detecting the aspect categories is treated as a classification task and the target detection as a sequence labelling task, which is achieved using the CRF decoding or softmax decoding on the token representations. Xue et al. (2017) used a CNN and LSTM to jointly detect the targets and aspect categories in a sentence, while Wan et al. (2020) used a joint modelling objective by fine-tuning a pre-trained language model BERT to obtain the token representation.

TASD The TASD task aims to identify the triple: target, aspect categories, and polarities together. This is also an under explored task with a few past works. Brun and Nikoulina (2018) relied on available parsers and domain-specific semantic lexicons, but this method performs poorly as shown in our experiments. Another method is the TAS-BERT joint modeling method that fine-tunes the pre-trained model BERT to solve the aspect-sentiment detection task using the classification token and the detecting the targets corresponding to ASD using the token classification with CRF/softmax decoding.

All the above methods use a classification or sequence labeling method to solve the sub-tasks of ABSA. However, we employ a conditional text generation based method to solve the tasks simultaneously to capture both implicit and explicit targets corresponding to an aspect-sentiment pair effectively.

6 Conclusion

In this study, we use an unique generative framework to handle diverse ABSA challenges. With a conditional text generation model, we tackle numerous ABSA sub-tasks, including many sentiment pair and triplet extraction tasks, by constructing the target sentences with our suggested pseudo-phrases and pseudo-sentences. To demonstrate the efficacy of our proposed technique, we conducted comprehensive testing on several benchmark datasets from various domains spanning six ABSA sub-tasks.

To demonstrate the efficacy of our formulation, we ran tests on a limited fraction of ABSA sub-tasks and variations. As a result, this could be a first step toward investigating and converting most of ABSA's classification tasks, as well as other text classification tasks outside of ABSA, into a conditional text generative framework, in order to reap the benefits of this formulation and a better understanding of the given text.

Acknowledgements

The authors acknowledge the use of the Sabine Cluster and the advanced support from the Research Computing Data Core at the University of Houston to carry out the research presented here.

References

- Abdulaziz Alghunaim. 2015. *A vector space approach for aspect-based sentiment analysis*. Ph.D. thesis, Massachusetts Institute of Technology.
- Caroline Brun and Vassilina Nikoulina. 2018. <https://doi.org/10.18653/v1/W18-6217> Aspect based sentiment analysis into the wild. pages 116–122.
- J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.
- Mengting Hu, Shiwan Zhao, Li Zhang, Keke Cai, Zhong Su, Renhong Cheng, and Xiaowei Shen. 2019a. <https://doi.org/10.18653/v1/D19-1467> CAN: Constrained attention networks for multi-aspect 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 4601–4610, Hong Kong, China. Association for Computational Linguistics.
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019b. <https://doi.org/10.18653/v1/P19-1051> Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.
- Niklas Jakob and Iryna Gurevych. 2010. <https://aclanthology.org/D10-1101> Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1035–1045, Cambridge, MA. Association for Computational Linguistics.
- Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. <https://doi.org/10.3115/v1/S14-2076> NRC-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442, Dublin, Ireland. Association for Computational Linguistics.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2019. <http://arxiv.org/abs/1910.13461> BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *CoRR*, abs/1910.13461.
- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2018a. <http://arxiv.org/abs/1811.05082> A unified model for opinion target extraction and target sentiment prediction. *CoRR*, abs/1811.05082.
- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. <http://arxiv.org/abs/1811.05082> A unified model for opinion target extraction and target sentiment prediction.
- Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018b. <http://arxiv.org/abs/1805.00760>

- Aspect term extraction with history attention and selective transformation.
- Xin Li and Wai Lam. 2017. <https://doi.org/10.18653/v1/D17-1310> Deep multi-task learning for aspect term extraction with memory interaction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2886–2892, Copenhagen, Denmark. Association for Computational Linguistics.
- Xinlong Li, Xingyu Fu, Guangluan Xu, Yang Yang, Jiuniu Wang, Li Jin, Qing Liu, and Tianyuan Xiang. 2020. <https://doi.org/10.1109/ACCESS.2020.2978511> Enhancing bert representation with context-aware embedding for aspect-based sentiment analysis. *IEEE Access*, PP:1–1.
- Bin Liang, Jiachen Du, Ruifeng Xu, Binyang Li, and Hejiao Huang. 2019. <https://doi.org/10.18653/v1/P19-1462> Context-aware embedding for targeted aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4678–4683, Florence, Italy. Association for Computational Linguistics.
- Fei Liu, Trevor Cohn, and Timothy Baldwin. 2018. <https://doi.org/10.18653/v1/N18-2045> Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 278–283, New Orleans, Louisiana. Association for Computational Linguistics.
- Kang Liu, Liheng Xu, and Jun Zhao. 2013. <https://aclanthology.org/P13-1172> Syntactic patterns versus word alignment: Extracting opinion targets from online reviews. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1754–1763, Sofia, Bulgaria. Association for Computational Linguistics.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. <http://arxiv.org/abs/1907.11692> Roberta: A robustly optimized bert pretraining approach.
- Huaishao Luo, Tianrui Li, Bing Liu, Bin Wang, and Herwig Unger. 2018. <http://arxiv.org/abs/1805.07889> Improving aspect term extraction with bidirectional dependency tree representation. *CoRR*, abs/1805.07889.
- Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang. 2019. <https://doi.org/10.18653/v1/P19-1056> DOER: Dual cross-shared RNN for aspect term-polarity co-extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 591–601, Florence, Italy. Association for Computational Linguistics.
- Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm.
- Margaret Mitchell, Jacqui Aguilar, Theresa Wilson, and Benjamin Van Durme. 2013. <https://aclanthology.org/D13-1171> Open domain targeted sentiment. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1643–1654, Seattle, Washington, USA. Association for Computational Linguistics.
- Sajad Movahedi, Erfan Ghadery, Heshaam Faili, and Azadeh Shakery. 2019. <http://arxiv.org/abs/1901.01183> Aspect category detection via topic-attention network.
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. <http://arxiv.org/abs/1910.10683> Exploring the limits of transfer learning with a unified text-to-text transformer.
- Meesala Shobha Rani and Sumathy Subramanian. 2020. Attention mechanism with gated recurrent unit using convolutional neural network for aspect level opinion mining. *Arabian Journal for Science and Engineering*, pages 1–13.
- Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. <https://doi.org/10.18653/v1/D15-1044>

- A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal. Association for Computational Linguistics.
- Martin Schmitt, Simon Steinheber, Konrad Schreiber, and Benjamin Roth. 2018. <https://doi.org/10.18653/v1/D18-1139> Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1109–1114, Brussels, Belgium. Association for Computational Linguistics.
- Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. *arXiv preprint arXiv:1903.09588*.
- Hai Wan, Y. Yang, Jianfeng Du, Y. Liu, Kunxun Qi, and J. Pan. 2020. Target-aspect-sentiment joint detection for aspect-based sentiment analysis. In *AAAI*.
- 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, AAAI'17*, page 3316–3322. AAAI Press.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. <https://doi.org/10.18653/v1/D16-1058> Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.
- Zhengxuan Wu and Desmond C. Ong. 2020. <http://arxiv.org/abs/2010.07523> Context-guided bert for targeted aspect-based sentiment analysis.
- Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2018. <https://doi.org/10.18653/v1/P18-2094> Double embeddings and CNN-based sequence labeling for aspect extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 592–598, Melbourne, Australia. Association for Computational Linguistics.
- Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2019. <http://arxiv.org/abs/1904.02232> Bert post-training for review reading comprehension and aspect-based sentiment analysis.
- Wei Xue and Tao Li. 2018. <https://doi.org/10.18653/v1/P18-1234> Aspect based sentiment analysis with gated convolutional networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2514–2523, Melbourne, Australia. Association for Computational Linguistics.
- Wei Xue, Wubai Zhou, Tao Li, and Qing Wang. 2017. <https://www.aclweb.org/anthology/I17-2026> MTNA: A neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 151–156, Taipei, Taiwan. Asian Federation of Natural Language Processing.
- Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. <http://arxiv.org/abs/1605.07843> Unsupervised word and dependency path embeddings for aspect term extraction. *CoRR*, abs/1605.07843.
- Jiangfeng Zeng, Xiao Ma, and Ke Zhou. 2019. <https://doi.org/10.1109/ACCESS.2019.2893806> Enhancing attention-based lstm with position context for aspect-level sentiment classification. *IEEE Access*, 7:20462–20471.
- Meishan Zhang, Yue Zhang, and Duy-Tin Vo. 2015. <https://doi.org/10.18653/v1/D15-1073> Neural networks for open domain targeted sentiment. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 612–621, Lisbon, Portugal. Association for Computational Linguistics.