Generative Opinion Triplet Extraction Using Pretrained Language Model

William
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
13518138@std.stei.itb.ac.id

Masayu Leylia Khodra
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
masayu@itb.ac.id

***Abstract*— Previous researches on opinion triplet extraction for Indonesian reviews in the hotel domain has been conducted in a discriminative manner using a sequence labeling approach. However, the opinion triplet extracted by those research is limited to explicit opinion triplet only, while neglecting the opinion triplets that contains implicit aspects. In this paper, we build a model that can perform opinion triplet extraction with explicit and implicit aspects based on the seq2seq approach. The system can extract opinion triplets using pre-trained language models that have been fine-tuned on datasets with label in extraction-style paradigm. The generated text then can be extracted to get the opinion triplets. By transforming opinion triplet extraction into a text generation problem with the help of a pre-trained language model (IndoT5), we are able to achieve a significant improvement of 4% on the F1 score compared to the result from previous researches.**

Keywords*— *opinion triplet extraction; text generation; pre-trained language model

I. Introduction

Sentiment analysis has been widely used among companies to automatically extract opinions about their products or services. Several companies may need more fine-grained analysis by employing aspect-based sentiment analysis (ABSA). It is usually formulated as detecting aspect terms and sentiments expressed in a sentence towards the aspects (Hu et al., 2019). Opinion triplet extraction is a task that attempts to extract (aspect term, opinion term, polarity sentiment) triplets from the given sentence, which tells what the opinion target (aspect) is, how its sentiment orientation (polarity) is, and why such review have that sentiment orientation expressed through the opinion term (Peng et al., 2019). Therefore, a model extracting opinion triplets shows more complete sentiment information, compared with previous models working for individual tasks.

However, current studies on aspect-based sentiment analysis for Indonesian reviews in the hotel domain such as in Fernando et al. (2019) and Genadi and Khodra (2022) only considered the extraction of explicit aspects, while ignoring the implicit ones. A recent state-of-the-art approach for opinion triplet extraction for Indonesian review text in the hotel domain adopts the Grid Tagging Scheme from Wu et al. (2020), which labels each token pair from the input text with tags and then matches those tags to get opinion triplets from input text also has the same limitation, that is only able to extract explicit triplet opinions. Product reviews contain many implicit aspects and opinions (Cai et al., 2021). Moreover, in most existing ABSA benchmark datasets, the given text is a single sentence where the entity is assumed to be known (Zhang et al., 2022). Meanwhile, the lack of studies that discusses and solves the implicit aspect problems in opinion triplet extraction for Indonesian reviews in the hotel domain motivates us to build a model that can handle current state-of-the-art approach limitation for Indonesian review in the hotel domain.

To circumvent the above issues, we adopt the text generation approach using the pre-trained language model proposed by Zhang et al. (2021) to generate text containing opinion triplets from input text. This is done to alleviate issues on unified tagging scheme when handling implicit aspects by using the characteristics of language models that were able to generate text that were not not limited to vocabulary from input text only. An extraction-style paradigm is employed the dataset labels. Given a sentence, the labels would be a text containing triplet opinions 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.

The contributions of this work can be summarized as follows. 1) We construct new datasets for the opinion triplet extraction task that include implicit aspects for Indonesian reviews in the hotel domain based on Genadi and Khodra (2022) dataset. 2) We show that the usage of text generation in opinion triplet extraction tasks for Indonesian reviews can be done, by changing the label of the dataset using the extraction-style paradigm proposed by Zhang et al. (2021) and the usage of post-process to fix the generated text. 3) We propose model variations from the one proposed by Zhang et al. (2021) by conducting post-training on the pre-trained language models (IndoT5), introducing a new normalization strategy in post-process, and varying model beam-width during inference. 4) We prove that proposed model outperforms the previously best-reported results in opinion triplet extraction task for Indonesian reviews.

The remaining aspect of this research is organized as follow: section 2 discusses related works i.e. Generative Aspect Based Sentiment Analysis, T5, and Grid Tagging Scheme. Section 3 describe proposed system. The experiment setup and dataset used for this work will be described in section 4. Section 5 will explain the results and the analysis of the experiment. Finally, the conclusions of this work will be discussed in section 6.

II. Related Work

A. *Generative Aspect Based Sentiment Analysis (GAS)*

Research related to aspect-based sentiment analysis generally requires a specific model development process to adjust the inputs and outputs of each task in aspect-based sentiment analysis. Zhang et al., (2021) propose a Generative Aspect-based Sentiment Analysis (GAS) which aims to form a general approach that can solve various aspects of aspect-based sentiment analysis with a text generation approach including opinion triplet extraction task.

To transform opinion triplet extraction task, which is generally carried out using a sequence labeling or text classification approach into a text generation approach, Zhang et al., (2021) propose two annotation paradigms for annotating training data. The two annotation paradigms are called annotation-style and extraction-style. The main difference between the two paradigms can be seen in the annotation of the triplet opinion pair from the review sentence “The Unibody construction is solid, sleek and beautiful”. In the extraction-style model, the output target would be a natural language consisting of triplet opinions from that review text, namely “(Unibody construction, solid, positive); (Unibody construction, sleek, positive); (Unibody construction, beautiful, positive)”, while in the annotation-style model the output target is the input text along with additional information in the form of an opinion triplet, namely “The [Unibody construction | positive | solid, sleek, beautiful] is solid, sleek and beautiful”.

The corpus is then used to train the T5 pre-trained language model (Raffel et al., 2020) as a language model to generate opinion triplets. After the training process is carried out, the desired triplet of opinion can be obtained by extracting the output of the model (y') according to the annotation paradigm used in the training data. For example, for extraction-style, the content included in brackets "()" were extracted from y'. Each of the different opinion triplet elements (a, o, p) can be separated with a comma "," as a separator. If the extraction process fails due to a format error in y' then the prediction is ignored.

Also, to handle the issue when the generated sentiment element doesn't exist on input text Zhang et al (2021) proposed a prediction normalization step. 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 Vc. For aspect term and opinion term, Vc contains all words in the current input sentence x; for aspect category, Vc is a collection of all categories in the dataset; for sentiment polarity, Vc 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 Vc, we use e ∈ Vc, which has the smallest Levenshtein distance with e, to replace e.

B. *Text-to-Text Transfer Transformers (T5)*

The T5 model was introduced by Raffel et al. (2021). This language modeling approach is based on two phases: pre-training, which allows defining a shared knowledge-base useful for a large class of sequence-to-sequence tasks, and fine-tuning, which specializes the model to specific tasks of interest. The T5 model is based on the transformer model architecture that allows it to handle a variable-sized input using stacks of self-attention layers instead of RNNs or CNNs. When an input sequence is provided, it is mapped to a sequence of embeddings that are passed into the encoder. The encoders are all identical in structure and each one is composed of two subcomponents: a self-attention layer followed by a small feed-forward network. Normalization layer is applied to the input of each subcomponent while a residual skip connection adds each input of the subcomponent to its output. Meanwhile, dropout layer is applied within the feed-forward network, on the skip connection, on the attention weights, and at the input and output of the entire stack. The decoders work similarly to the encoders: Each self-attention layer is followed by an additional attention mechanism that attends to the output of the encoder. The output of the final decoder block is fed into a dense layer with a softmax output, to produce the output probabilities over the vocabulary.

T5 offer two main advantages over other state-of-the-art models: (i) it is more efficient than RNNs since it allows for computing the output layers in parallel, and (ii) it can detect hidden and long-ranged dependencies among tokens, without assuming that nearest tokens are more related than distant ones. The available T5 pre-training language model for Indonesian is IndoT5. The model was trained on an Indonesian-language corpus from the filtered mC4 (Xue, et al., 2021) corpus.

C. *Grid Tagging Scheme (GTS)*

Wirawan (2021) developed an opinion triplet extraction system by adopting the Grid Tagging Scheme (GTS) approach from Wu et al., 2020 for Indonesian reviews in the hotel domain. When tackling the opinion triplet extraction task, Grid Tagging Scheme (GTS) uses sets of tags {A, O, Pos, Neu, Neg, N} to denote the relation of word-pair (wi, wj) in a sentence. The three tags Pos, Neu, and Neg respectively indicate positive, neutral, or negative sentiment expressed in the opinion triplet consisting of the word-pair (wi, wj) and also used to denote the relation between aspect term and opinion term. A tagging example of the extraction triplet opinion task can be seen in Fig. 1.

The	hot	dogs	are	top	notch	but	average	coffee	
N	N	N	N	N	N	N	N	N	The
	A	A	N	Pos	Pos	N	N	N	hot
		A	N	Pos	Pos	N	N	N	dogs
			N	N	N	N	N	N	are
				O	O	N	N	N	top
					O	N	N	N	notch
						N	N	N	but
							O	Neu	average
								A	coffee

Fig. 1. Example of Tagging Result from GTS (Wu et al., 2020)

III. Methodology

Our system is designed based on Zhang et al. (2021) proposed solution to solve the opinion triplet extraction task. Instead of formulating the opinion triplet extraction task as a sequence tagging problem, we use a text generation approach. Given the input sentence x, we generate a target sequence y' using the extraction-style paradigm, with a text generation model f(·). Then the desired sentiment pairs or triplets can be decoded from the generated sequence y'. Because we use the extraction-style modeling, we can separate the generated triplet opinion from the sequence y' and ignore those invalid generations. The goal is to extract all triplet opinions from the sentence by generating text containing that triplet opinion from a given sentence.

Also currently, the available dataset (Indonesian reviews in the hotel domain) was annotated with only explicit triplet opinion. We further added annotation on triplet opinion with implicit aspects. The annotation is done simply by pairing the expression of sentiment without an aspect pair with `NULL` and adding information about the sentiment polarity (positive, negative, neutral) on the opinion triplets with the implicit aspect. An illustrative example of annotation done in this study can be seen in TABLE I.

TABLE I. ANNOTATION EXAMPLE ON OPINION TRIPLET

Text Review	
oke banget, tetapi ac nya tidak bisa diatur suhunya. (*really okay, but the AC temperature can't be set.*)	
Triplet Opinion (Explicit)	**Triplet Opinions (Implicit Aspect)**
[(ac nya, tidak bisa diatur, **NEG)**] *[(AC, can't be set,* **NEG***)]*	[(`NULL`, oke banget, **POS**)] *[(`NULL`, really okay,* **POS***)]*

A. IndoT5 as Generation Model

We use Text-to-Text Transfer Transformer (T5) (Raffel et al., 2020), a Transformer based architecture that uses a text-to-text approach. It achieves state-of-the-art performances across a variety of NLP tasks. Specifically, we use IndoT5-base, a T5-base model trained on the Indonesian dataset, as our generation model.

We perform lowercase and tokenization of the input sentence using a 32,000 word-piece vocabulary with the Sentencepiece tokenizer for the preprocessing step. Then we add padding and truncation towards 128 sequences to ensure the model input has the same size. The fine-tune objective for the model takes an input sequence X = {x1, ..., xn} and aims to generate an output sequence Y = {y1, ..., ym} that contains an opinion triplet from inputs in extraction-style paradigm. Here we treat the desired opinion triplets as the target, which resembles the direct extraction of the expected sentiment elements in a generative manner.

Also, we perform post-training to train further the pre-trained model. Based on Aditya and Khodra (2022), post-training is used to further the pre-training process of language models such as BERT, GPT, and T5 to increase model knowledge in certain domains. To do this, we use unlabeled hotel domain data consisting of 136,788 sentences from Aditya and Khodra's (2022) research to increase pre-train model understanding of the hotel domain terminology text in general.

B. Opinion Triplet Decoding

Upon obtaining the generated text, we conduct a triplet decoding process. We separate the generated triplets from the generated sequence and extract the contents included in the bracket "()". Then, we get the aspect term, sentiment term, and polarity sentiment from each triplet by separating them based on the "," delimiter. If such decoding fails due to the output sequence not following the extraction-style paradigm, we ignore such predictions.

For example, from word sequence "kamarnya bagus, tetapi toilet kurang bersih."(the room is good, but the toilet is not clean.), we get generated text "(kamarnya, bagus, positif); (toilet, kurang bersih, negatif)". The opinion triplet then can be extracted: [(kamarnya, bagus, **POS**), (toilet, kurang bersih, **NEG**)] (*translation:* [(*the room, good,* **POS)**, (*toilet, not clean,* **NEG**)]).

C. Post-process

We use post-process to check and correct the results of triplet generation that do not exist in the input text. While ideally, the generated element e ∈ s after decoding is supposed to exactly belong to the vocabulary set it is meant to be, this might not always hold since each element is generated from the vocabulary set containing all tokens instead of its specific vocabulary set (Zhang et al., 2021).

To combat this we propose a post-process similar to the prediction normalization strategy proposed by Zhang et al. (2021), if the predicted terms are not in the input text, it will try to look for the most similar vocabulary from the input text to replace the predicted terms. In order to find the most similar term from the input text, we use similarity function.

A similarity function or measure is a real-valued function that quantifies the similarity between two objects. In this study, we present two similarity measures: Levenshtein distance and cosine similarity. Those two measures will be used as similarity functions during post-process. Illustrations for both similarity measures usage in post-process can be seen in Fig. 2.

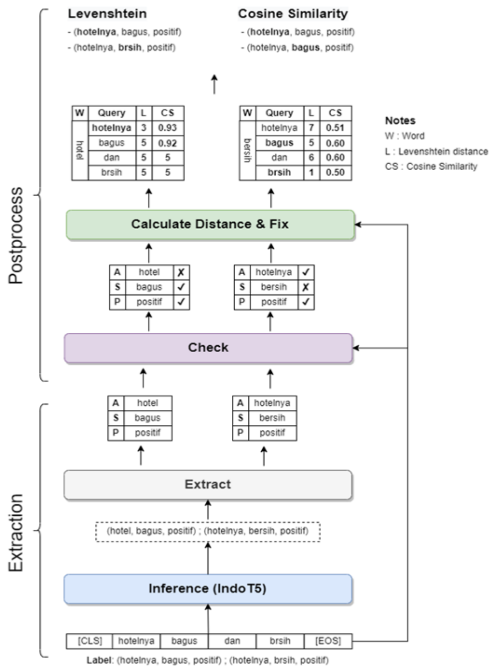

Fig. 2. Illustration of post-process using Levenshtein distance and cosine similarity

Levenshtein Distance. A similarity measure between words. Given two words, the distance measures the number of edits needed to transform one word into another (insertion, deletion, and substitution). It is used by Zhang et al. (2021) as a simple yet effective strategy to alleviate issues where the morphology shifts where two words have minor lexical differences or when the model might generate words with the same etyma but different word types.

Cosine Similarity. A similarity measure that measures the cosine of the angle between two vectors (text representation) to handle error cases when the model generates words that are quite different lexically from the target but still have similar meanings.

Additionally, there is a difference in post-process for the handling of opinion triplet extraction with implicit aspects. If the predicted aspect does not exist in the vocabulary of the input text but is a special aspect (`NULL`), we ignore those prediction.

IV. EXPERIMENT SETUP

A. Datasets and Evaluation Metrics

Dataset. We conduct experiments on two datasets in the "hotel" domain with a total of 5000 instances. The statistics of these datasets are shown in TABLE II. Furthermore, these two datasets will be split into three subsets, namely the training set (3000 instances), validation set (1000 instances), and test set (1000 instances) same as used in previous studies (Wirawan, 2021) so we can compare the model performance equally. Also, the difference between these two datasets is whether we include triplet opinion with implicit aspects or not. The first dataset, we call "EXPLICIT Dataset" does not include triplet opinion with implicit aspects, while the second dataset, we call "COMBINED Dataset" contains triplet opinion with implicit aspects that we annotated earlier.

Evaluation Metrics. We adopt precision, recall, and f1-score as evaluation metrics for opinion triplet extraction task. Only exactly matched triplet, i.e., with all of the aspects, opinions, and sentiment matched against the gold label, are viewed as correct (true positive) during evaluation. The McNemar test is used to examine statistical significance of the results when comparing proposed solution performance with baseline.

TABLE II. DATASET STATISTICS

Dataset		# sentences	# triplets (explicit)	# triplets w/implicit	# total
EXPLICIT	train	3000	6251	0	6251
	val	1000	1989	0	1989
	test	1000	2291	0	2291
COMBINED	train	3000	6251	1290	7541
	val	1000	1989	467	2456
	test	1000	2291	418	2709

B. Baseline and Variants

We compare the proposed text generation approach with the current state-of-the-art sequence tagging approach for Indonesian reviews in the hotel domain - Grid Tagging Scheme (Wu et al., 2020) i.e., for each word pair, the model predicts whether they belong to the same aspect, the same opinion, relation between aspect and opinion pair (denoted by polarity sentiment), or none of the above that was adopted by Wirawan (2021) for Indonesian reviews. The comparison will be done for the opinion triplet extraction task with explicit aspects using the EXPLICIT dataset.

We also propose a list of variants of our proposed text generation approach to examine the efficacy of different components in it: (1) **post-training**, where we compare the performance when using the IndoT5 model or IndoT5 after post-training to see the effect of post-training on the opinion triplet extraction task using text generation (we call the post-trained model IndoT5-IDPT), (2) **post-process**, where we compare performance for different similarity measures, such as Levenshtein distance and cosine similarity, and (3) **beam-width**, where we compare different beam size during model inference. For all model variants, we use similar experimental settings for simplicity: we train the model with the batch size of 8 and. The learning rate is set to be 3e-4 and the model is trained up to 10 epochs.

V. RESULT AND ANALYSIS

Variants Comparison. First, we compare each variant of our proposed solution. The comparison with variants of our proposed solution (GAS) on validation data can be shown in TABLE IV. TABLE V. TABLE VI. .

Post-training. When we compare the usage of IndoT5 vs IndoT5-IDPT, we found out that IndoT5 alone perform better than its post-trained counterpart. Rather than performing opinion triplet extraction, we notice that IndoT5-IDPT is more sensitive towards sentiment expression than IndoT5, which causes the model to often extract the opinion triplet pairs incorrectly compared to IndoT5. For example, in the text "nyaman dan bersih." (*comfortable and clean*) IndoT5 will not predict any triplets, but IndoT5-IDPT will predict "(kamar (*room*), nyaman (*comfortable*), **POS**), (kamar, bersih (*clean*), **POS**)". While it is shown that the model is more capable of understanding implicit aspects after post-training, it is reasonable that the model gets a lower score on the evaluation metric because of exact matching.

Post-process. When we compare the usage of raw (no post-process) vs Levenshtein distance vs cosine similarity, we found out that the usage of post-process increases the proposed solution performance. Specifically, the usage of post-process with Levenshtein distance is slightly better than other post-process. We notice that there are a lot of typo terms in the dataset and the post-process with Levenshtein distance can handle those cases slightly better than other proposed post-process in this study.

Beam-width. Lastly, we compare the beam width during model inference. While increasing the beam size improves the performance in Raffel et al., (2021) for tasks with longer output sequences, we notice that the text in our dataset mostly consists of short text. Thus, this might be the reason why there are no significant improvements after increasing the beam width.

TABLE III. CASE STUDY.

Case	Ground Truth (Label)	Proposed Model (EXPLICIT ONLY)	Proposed Model (COMBINED)
bagus dan bersih , namun pelayan cerewet . (*nice and clean , but the waiter was talkative.)*	[(pelayan, cerewet, **NEG**), (NULL, bagus, **POS**) †, (NULL, bersih, **POS**) †] *[(waiter, talkative,* **NEG**), *(NULL, nice,* **POS***)* †, *(NULL, clean,* **POS***)* †*]*	[(pelayan, cerewet, **NEG**)] [(*waiter, talkative*, **NEG**)]	[(pelayan, cerewet, **NEG**), (NULL, bagus, **POS**), (NULL, bersih, **POS**)] [(*waiter, talkative*, **NEG**), (*NULL, nice*, **POS**) , (*NULL, clean*, **POS**)]
pokok nya puas , tenang suasana nya . (satisfied, the atmosphere is calm.)	[(suasana, tenang, **POS**), (NULL, puas, **POS**) †] *[(atmosphere, calm,* **POS***),* *(NULL, satisfied,* **POS***)* †*]*	[(suasana, tenang, **POS**) (suasana [X], puas, **POS**)] *[(atmosphere, calm,* **POS***),* *(atmosphere* [X]*, satisfied,* **POS***)]*	[(suasana, tenang, **POS**) (NULL, puas, **POS**)] *[(atmosphere, calm,* **POS***),* *(NULL, satisfied,* **POS***)]*
menyenangkan menginap disini . (it's enjoyable staying here)	[(menginap, menyenangkan, **POS**)] *[(staying, enjoyable,* **POS***)]*	[()[X]]	[()[X]]
tidak terlalu jauh dari pusat kota . (it's not too far from the city center)	[(dari pusat kota, tidak terlalu jauh, **POS**)] *[(from city center, not too far,* **POS***)]*	[()[X]]	[(NULL [X], tidak terlalu jauh, **POS**)] *[(NULL* [X]*, not too far,* **POS***)]*

a. Marker *X* indicate incorrect predictions

b. Marker † denotes results for implicit aspects

TABLE IV. PRETRAINED MODEL VARIANT COMPARISON

Pretrained Model	EXPLICIT			COMBINED		
	P	**R**	**F1**	**P**	**R**	**F1**
IndoT5	**0.812**	**0.848**	**0.829**	**0.806**	**0.860**	**0.832**
IndoT5-IDPT	0.702	0.789	0.743	0.792	0.852	0.821

TABLE V. POST-PROCESS VARIANT COMPARISON

Postprocess	EXPLICIT			COMBINED		
	P	**R**	**P**	**R**	**P**	**R**
Raw	0.812	0.848	0.829	0.806	0.860	0.832
Levenshtein	**0.817**	**0.854**	**0.835**	**0.809**	**0.863**	**0.835**
Cosine Sim	**0.817**	0.853	0.835	0.808	**0.863**	**0.835**

TABLE VI. BEAM-WIDTH VARIANT COMPARISON

Beam-width	EXPLICIT			COMBINED		
	P	**R**	**P**	**R**	**P**	**R**
k=1	**0.812**	**0.848**	**0.829**	**0.806**	**0.860**	**0.832**
k=5	0.812	0.848	0.829	0.806	0.860	0.832

Baseline Comparison. The results in comparison with baselines are shown in TABLE VII. using test data, but only for datasets without triplet opinion with implicit for a fair comparison. Our proposed model OTE-MTL outperforms the sequence tagging approach for Indonesian reviews in the hotel domain on explicit datasets. We use McNemar test with a significant value of 5% ($p < 0.05$). The test was carried out using a one-tailed test to determine whether the performance of the TA model was better than the GTS-GW model. Therefore, the null hypothesis (H0) from the statistical test is that both models have the same performance (same error rate). From the results of the McNemar statistical test scores, the McNemar statistical value ($\chi 2$) was 95.5 with a p-value of 1.49 x 10-22 so we can reject the null hypothesis (H0) which states that both models have the same performance. Thus, we conclude our proposed solution is more effective in dealing with opinion triplet extraction tasks.

TABLE VII. MAIN RESULT OF OPINION TRIPLET EXTRACTION TASK

Model	EXPLICIT			COMBINED		
	P	**R**	**P**	**R**	**P**	**R**
GTS	**0.857**	0.703	0.773	-	-	-
GAS (ours)	0.851	**0.775**	**0.811**	**0.820**	**0.771**	**0.795**

Error Analysis. To understand the limitation of our proposed solution in detail we conduct a detailed analysis of false positives (extracted by the system but not existing in ground truth) and false negatives (not extracted by the system but existing in ground truth) on both test datasets. The example case can be observed from TABLE III.

Based on the analysis, we found out that our proposed solution often experiences prediction errors when there is a sentiment expression without an aspect pair. Not only that, our model often experiences a prediction failure on the opinion triplet with aspect expressions in the form of adjectives.

Also the addition of implicit aspects also makes the model often predict aspects that are rarely found in the training data to be predicted as special aspects (NULL). In addition, the circumstance might also reflect that an exact match evaluation metric is not an ideal metric when systems are evaluated, since minor discrepancies in a span may be harmless for opinion interpretation in practice.

VI. CONCLUSION

In this paper, we solve the opinion triplet extraction task for Indonesian reviews in the hotel domain by formulating the target sentences with extraction-style paradigms based on work done by Zhang, et al. (2021). Handling of opinion triplet pairs with implicit aspects can be done by pairing sentiment expressions without pairs with a special aspect (NULL) based on annotation done by Cai et al. (2021). We also compare evaluation results with the result from the previous study and analyze error cases that still occur in the study. This work is an initial attempt on transforming ABTE tasks, which are typically treated as sequence labeling problems, into text generation problems. For future works, we suggest using a variation of another pre-trained language model with a larger number of parameters such as IndoT5-large to improve the performance of the opinion triplet extraction model. Research by Raffel et al., (2020) proves that increasing the number of model parameters can improve model performance. Also, following this direction, designing more effective generation paradigms and post-process and extending such ideas to other tasks.

REFERENCES

[1] Genadi, R. A., & Khodra, M. L. (2022). Opinion Triplet Extraction for Aspect-Based Sentiment Analysis Using Co-Extraction Approach. Journal of Information and Communication Technology, 21(2), 255-277.

[2] J. Fernando, M. L. Khodra and A. A. Septiandri, "Aspect and Opinion Terms Extraction Using Double Embeddings and Attention Mechanism for Indonesian Hotel Reviews," 2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2019, pp. 1-6.

[3] Aditya, I. P. E. S. (2022). Post-Training for Aspect-based Sentiment Analysis in Indonesian Language. Graduate Thesis, Institut Teknologi Bandung, Bandung

[4] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, & Peter J.Liu(2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21(140), 1-67.

[5] 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 Annual Meeting of the Association for Computational Linguistics, pages 537–546, Florence, Italy. Association for Computational Linguistics.

[6] Peng, Haiyun & Xu, Lu & Bing, Lidong & Huang, Fei & Lu, Wei & Si, Luo. (2020). Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis. Proceedings of the AAAI Conference on Artificial Intelligence. 34. 8600-8607. 10.1609/aaai.v34i05.6383.

[7] Sitikhu, Pinky & Pahi, Kritish & Thapa, Pujan & Shakya, Subarna. (2019). A Comparison of Semantic Similarity Methods for Maximum Human Interpretability

[8] Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing and Wai Lam (2021). Towards Generative Aspect-Based Sentiment Analysis.

[9] Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing and Wai Lam (2022). A Survey on Aspect Based Sentiment Analysis: Tasks, Methods, and Challenges.

[10] Wirawan, G. P. (2021). Ekstraksi Triplet Opini Pada Analisis Sentimen Berbasis Aspek dengan Pendekatan Grid Tagging Scheme. Undergraduate Thesis, Institut Teknologi Bandung, Bandung.

[11] Xue, C. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 483–498). Association for Computational Linguistics.

[12] Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 2576–2585, Online. Association for Computational Linguistics.