International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Optimizing Aspect Term Extraction and Sentiment Classification through Attention Mechanism and Sparse Attention Techniques

Dhendra Marutho[1,2] Muljono[1]* Supriadi Rustad[1] Purwanto[1]

[1] *Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia*
[2] *Department of Informatics, Universitas Muhammadiyah Semarang, Semarang, Indonesia*
* Corresponding author's Email: muljono@dsn.dinus.ac.id

Abstract: Aspect-based sentiment analysis (ABSA) has become an essential field in Natural Language Processing (NLP) in recent years. ABSA not only categorizes sentiment as positive, negative, or neutral but also understands the specific aspects or topics discussed in the text review. This study focuses on two important elements of ABSA: Aspect Term Extraction (ATE) and Aspect Sentiment Classification (ASC). This study uses a combination of Sentence Embedding (SBERT) techniques, Part-of-Speech tagging, cosine-similarity calculation to assess words with their respective aspect labels, and the sparse attention mechanism (BIGBIRD) method, which has been proven to increase accuracy effectively and is effective in terms of time and memory usage. By applying this method to two hotel review datasets, Traveloka Review and Semeval 2016 dataset, it is proven to work well on two ABSA tasks, namely ATE and ASC. The results of the ATE test obtained an accuracy of 0.99, and the ASC test obtained an accuracy of 0.89. This study contributes to the advancement of ABSA by introducing a new methodology that improves the accuracy of aspect term extraction and sentiment classification. Additionally, it identifies avenues for future research, including exploring additional techniques to improve model performance and address potential limitations.

Keywords: Aspect-based sentiment analysis, Attention mechanism, Sparse attention mechanism, BERT, BigBIRD, POS tagger, Semantic similarity, Cosine similarity.

1. Introduction

In the era of rapidly developing information, sentiment analysis is becoming increasingly important in understanding public opinion and views on various topics, products, or services. In this context, Aspect-Based Sentiment Analysis (ABSA) emerges as a method that can dig deeper into opinions and evaluations of specific aspects of a subject [1].

ABSA is not only related to positive, negative, or neutral review sentiments, ABSA can recognize sentences that have aspects [2]. This provides deeper insight to decision-makers in various fields, such as marketing, customer service, and product analysis [3].

ABSA has four stages, namely Aspect Term Extraction (ATE), Aspect Category Detection (ACD), Opinion Term Extraction (OTE), and Aspect Sentiment Classification (ASC), which play an important role in understanding specific aspects of a sentence [4]. ATE aims to identify relevant aspects of the text, such as "product", "seller", and "expedition service", in the sentence "The product is excellent, the seller is friendly and responsive, but unfortunately the delivery is long". ACD groups into broader categories such as "product quality", "customer service", and "delivery".

Opinion Term Extraction (OTE) is responsible for identifying opinions from each aspect that has been extracted, such as positive sentiment for the product and seller, negative sentiment towards the expedition service because of the long delivery time. Finally, ASC classifies the sentiment of each aspect as positive, negative, or neutral, thus providing a more holistic understanding of the evaluation and opinion in the text.

Research related to ATE was conducted [5] using customer review datasets of Uber, TripAdvisor and Amazon. By testing the effectiveness of using machine learning CRD, SVM, and deep learning approaches CNN and LSTM where the deep learning approach is superior to machine learning. ATE and ASC using CNN were conducted by [6] with product review datasets, this study used double embedding, where the first domain is a pre-trained specific domain and the second is a general domain. This combination is very effective for aspect extraction.

If research [5, 6] is domain-specific, while research [7] uses multi-domain from various reviews to perform ATE and ASC, there are still shortcomings because the data is relatively small, so it does not reflect the actual performance. The use of the BERT model can overcome this because BERT is trained on a very large and diverse dataset before being applied to a specific task. This model has learned many common language patterns and structures during the pre-training phase. BERT already understands language well, so it requires less specific data to train it further.

Research [8] proposed BERT and CRF methods that utilize the hidden layers of the BERT model to produce deeper semantic representations of the input sequence, while the CRF task is the joint distribution of sequence labels for more accurate predictions, for the datasets used are semeval 2014 and semeval 2016.

The use of transformers methods for ATE and ASC research was carried out by [9, 10] on hotel reviews, in research [9] proposed TF-ICF to extract terms from reviews, LDA (Latent Dirichlet Allocation) to reveal hidden topics from each term while BERT to categorize sentiment aspects using semantic similarity. Research [10] uses a rule-based algorithm to obtain word types and relationships between sentences. This method aims to identify candidate aspects and opinions based on the type of sentence structure, For the ASC process using the BERT Embedding method and semantic similarity.

The use of grammar rules can also be done in the ATE and ASC processes as done by [11, 12] using grammar rules combined with other extraction techniques such as Elmo (Embeddings from Language Models) for contextual word representation, WordNet for synonyms and antonyms of words, TF-ICF (Term Frequency-Inverse Cluster Frequency) to assess the importance of a word in a document relative to a cluster, and semantic similarity to measure the closeness of meaning between words.

ABSA research on the ATE task is currently still dominated by the word embedding method in the feature extraction stage, such as research [9, 10, 13-15] This can be a problem because sentences lose meaning and are inaccurate. This study uses sentence embedding to overcome this problem. The method used is Sentence BERT (SBERT) combined with cosine similarity to extract aspects. The ATE and ASC classification methods used are the BIGBIRD Sparse Attention Mechanism, which only pays attention to important tokens, not all tokens [16]. The sparse attention mechanism makes classification faster than the full attention process using BERT [17]. The sparse attention mechanism makes attention calculations more efficient and memory efficient, but sparse attention has low accuracy compared to full attention. This low accuracy is a problem that must be solved, namely by combining the SBERT process to maintain meaning, combined with the Sparse attention mechanism to speed up the classification process and increase accuracy in the ABSA subtasks, namely ATE and ASC.

2. Related theory

Several theories related to the research are explained in this section.

2.1 Data preprocessing

Preprocessing a dataset is an essential step in data preparation before it is used for model training or evaluation. Preprocessing steps aim to clean and prepare the data to suit the needs of the analysis or modeling to be carried out. First, the raw data is cleaned from non-standard formats, such as inappropriate punctuation, or inconsistent formats. This step often involves cleaning the text by removing non-alphanumeric characters, converting the text to lowercase [18], and normalizing the text if necessary. The lemmatization process is continued because it is better at recognizing meaning, compared to stemming which only removes prefixes and suffixes [19].

2.2 Aspect Term Extraction (ATE)

ATE is an important stage in ABSA that aims to identify terms or words that represent certain aspects of entities in the text. For example, in a product review, ATE can identify the words "food", "service", and "price" in a restaurant review. Research [9, 10, 13-15], performs aspect extraction using several different approaches, research [9, 10, 13-15] uses topic modeling methods such as LDA, PLSA, and research [10] uses a rule-based method to obtain candidate aspects, then calculates aspects using cosine similarity, after obtaining aspects as, the classification process is carried out using SVM [15], LSTM [13, 14] dan BERT [9, 10].

Received: May 26, 2024. Revised: August 17, 2024.

2.3 Aspect Sentiment Classification (ASC)

ASC is an important stage in sentiment analysis that aims to classify sentiments associated with each aspect that has been extracted from a text. For example, after identifying aspects such as "food," "service," and "price" in a restaurant review, the next step is to determine the sentiment of each aspect, whether it is positive, negative, or neutral. Research [20-24] uses deep learning methods to classify sentiment aspects. this study will test the research using a combination of the SBERT method, part of speech tagger, and the BIGBIRD sparse attention mechanism transformers method.

2.4 Full attention mechanism BERT

Attention mechanisms have been a significant innovation in natural language processing (NLP), especially with the advent of the BERT (Bidirectional Encoder Representations from Transformers) model. BERT leverages attention mechanisms to understand the context of words in a sentence from both directions (left and right) [25], allowing it to capture complex semantic and syntactic nuances. Particularly in the context of ABSA [26], which includes both ATE and ASC, BERT's attention mechanism significantly improves identifying specific aspects and associated sentiments.

2.5 Full attention mechanism BERT

The sparse attention mechanism model BIGBIRD is a transformer model designed with the concept of sparse attention, offering an efficient solution to handle long-range relationships in text. This sparse attention mechanism limits the attention calculation to only a strategically selected subset of tokens, such as adjacent tokens or relevant tokens based on certain rules, thereby reducing the number of interactions between tokens in a sequence. Thus, the attention calculation becomes more efficient, and memory-saving compared to the traditional full attention mechanism [16].

On the ATE task, BIGBIRD can effectively identify aspect terms in the text by considering a wider context without sacrificing computational efficiency. This is important because aspects in sentences are often scattered and require a broad understanding of the context to be extracted accurately.

For the ASC task, BIGBIRD can classify sentiments associated with each aspect by considering long-range relationships in sentences that may affect sentiment polarity. The sparse attention mechanism allows the model to focus on important parts of the text that are relevant to a particular aspect, improving the accuracy of sentiment classification [17].

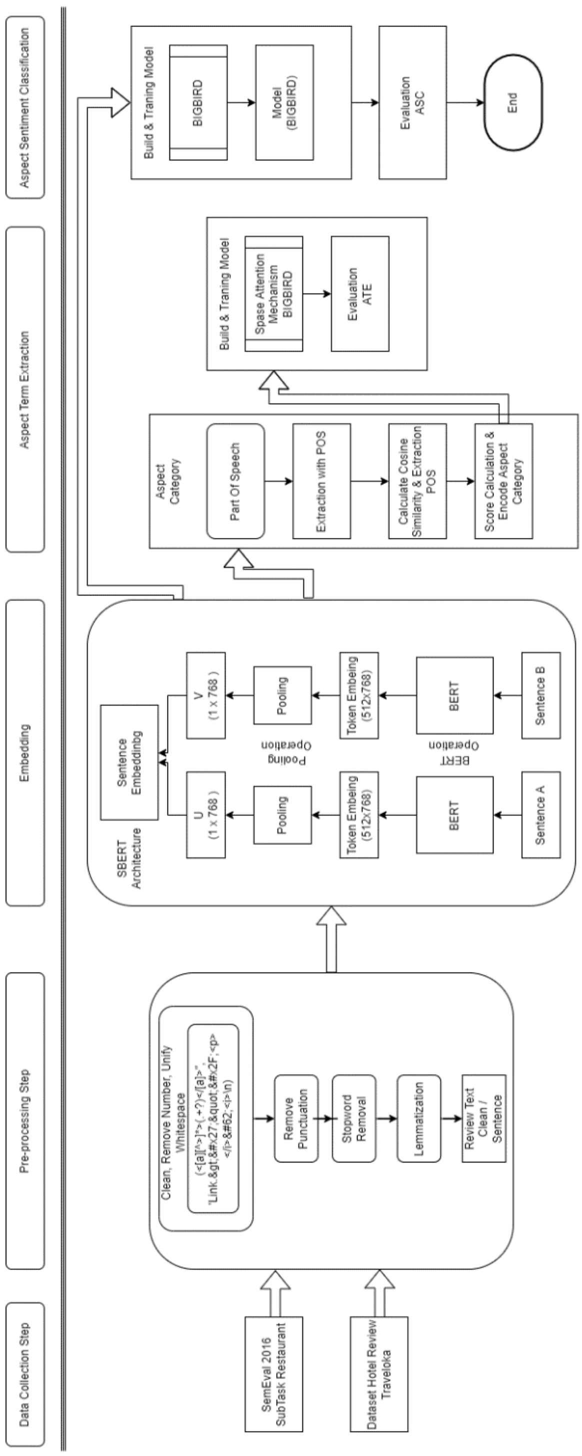

Figure. 1 Proposed Method

2.6 Semantic similarity

Semantic similarity is an important concept in natural language processing that measures the

Received: May 26, 2024. Revised: August 17, 2024.

closeness of meaning between words, phrases, or sentences. In the context of ABSA, understanding the semantic similarity between sentences or phrases is very important and is used to measure the meaning of each sentence [10].

Semantic similarity in the Sentence-BERT (SBERT) method is used to maintain semantic similarity after the BERT tokenization process. To measure the semantic similarity between two sentences generated by SBERT, cosine similarity is used. Cosine similarity is a method that measures the degree of similarity between two vectors in high-dimensional space by calculating the cosine of the angle between them. The cosine similarity formula is given by (1).

$$Cosine(S1, S2) = \frac{\sum_{i=1}^{k} S1i\ S2i}{\sqrt{\sum_{i=1}^{k} S1i^2}\sqrt{\sum_{i=1}^{k} S2i^2}} \quad (1)$$

Where S1 and S2 are the vectors to be compared, $S1_i$ dan $S2_i$ are the i elements of vectors S1 and S2, respectively, and k is the number of elements in the vectors. The cosine similarity value ranges between -1 and 1, where a value of 1 indicates that both vectors have the same direction (very similar). A value of 0 indicates that both vectors are perpendicular (not similar). A value of -1 indicates that both vectors have opposite directions (very dissimilar)

2.7 Evaluation

For the evaluation process, a confusion matrix is used to measure precision, recall, f1, and accuracy on ATE and ASC for evaluation formulas (2)-(5).

$$Precission = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - measure = \frac{2x\ Precission\ x\ recall}{Precission + recall} \quad (4)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

3. Research method

In this study, several datasets were tested, including the 2016 Semeval dataset for the restaurant subtask research [20-24, 27], and the hotel review dataset [9, 10, 13-15].

Table 1. Dataset Representation

Text
Judging from previous posts this used to be a good place, but not any longer.
We, there were four of us, arrived at noon - the place was empty - and the staff acted like we were imposing on them, and they were very rude.
They never brought us complimentary noodles, ignored repeated requests for sugar, and threw our dishes on the table.

Table 2. Cleaned Dataset

Clean Text
judging from previous posts this used to be a good place but not any longer.
we there were four of us arrived at noon the place was empty and the staff acted like we were imposing on them and they were very rude.
they never brought us complimentary noodles ignored repeated requests for sugar and threw our dishes on the table.

Table 3. Sentence Embedding

Sentence Embedding
['judging', 'from', 'previous', 'posts', 'this', 'used', 'to', 'be', 'a', 'good', 'place', 'but', 'not', 'any', 'longer']
['we', 'there', 'were', 'four', 'of', 'us', 'arrived', 'at', 'noon', 'the', 'place', 'was', 'empty', 'and', 'the', 'staff', 'acted', 'like', 'we', 'were', 'imposing', 'on', 'them', 'and', 'they', 'were', 'very', 'rude']
['they', 'never', 'brought', 'us', 'compliment', '##ary', 'noodles', 'ignored', 'repeated', 'requests', 'for', 'sugar', 'and', 'threw', 'our', 'dishes', 'on', 'the', 'table']

3.1 Preprocessing

The dataset used in this study cannot be used directly. It must go through preprocessing stages such as removing emojis, deleting hyperlinks, deleting punctuation (commas, periods, question marks), and changing words to lowercase. Sentence representation is shown in Table 1

The results of the dataset after the preprocessing stage are shown in Table 2.

Sentence tokenization stages using SBERT, the data is shown in Table 3 below:

3.2 Aspect extraction

The aspect extraction process of this study uses the research category aspects [9, 10, 13-15] in hotel reviews, for the category aspects are Cleanliness, Comfort, Food, Location, Service. This study

Received: May 26, 2024. Revised: August 17, 2024.

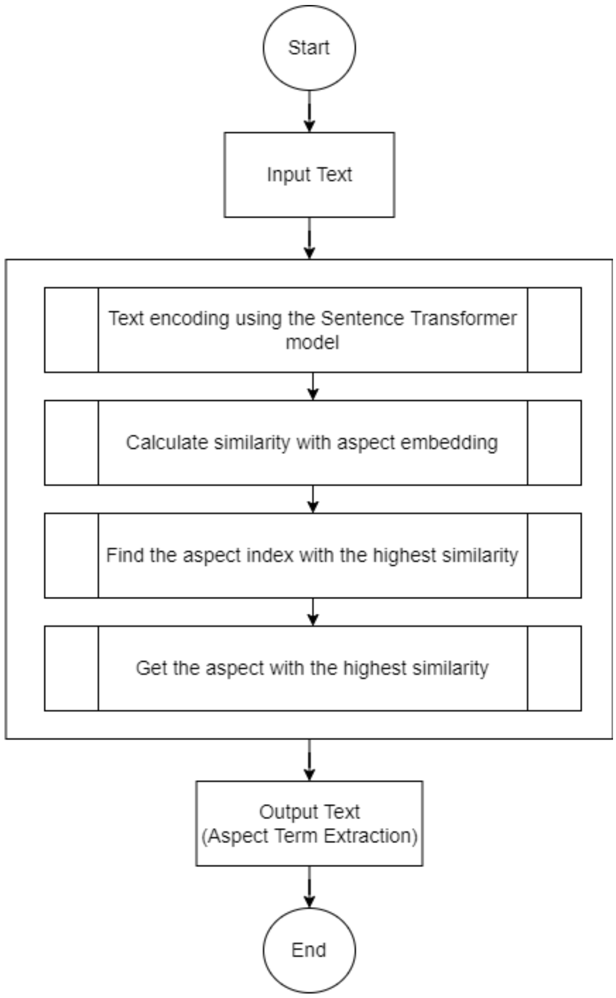

Figure. 2 Aspect extraction term process

Table 4. Aspect term extraction

Sentence Embeding	Aspect term extraction	Cosine Similarity Score
['judging', 'from', 'previous', 'posts', 'this', 'used', 'to', 'be', 'a', 'good', 'place', 'but', 'not', 'any', 'longer']	Location	0.984
['we', 'there', 'were', 'four', 'of', 'us', 'arrived', 'at', 'noon', 'the', 'place', 'was', 'empty', 'and', 'the', 'staff', 'acted', 'like', 'we', 'were', 'imposing', 'on', 'them', 'and', 'they', 'were', 'very', 'rude']	Service	0.991
['they', 'never', 'brought', 'us', 'compliment', '##ary', 'noodles', 'ignored', 'repeated', 'requests', 'for', 'sugar', 'and', 'threw', 'our', 'dishes', 'on', 'the', 'table']	Food	0.924

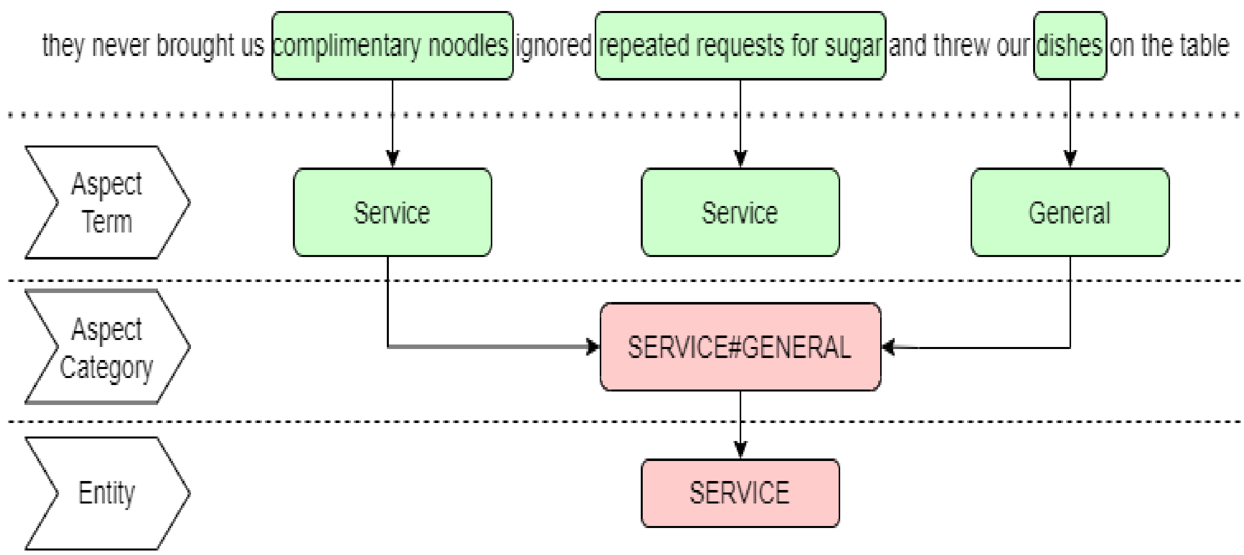

Figure. 3 Aspect Extraction Term Process

proposes a method using a combination of Sentence Transformers SBERT for sentence tokenization, calculating each semantic similarity of the category aspects from previous studies. The first step is to identify each sentence using part of speech tagging (POS), which involves separating each tag. Fig. 2 is a step in the form of a flowchart.

Explanation in Fig. 2.

1. The flow starts from the starting point.
2. The text is tokenized using the Sentence Transformer SBERT model.
3. The cosine similarity is calculated with the category aspect.
4. Find the aspect index with the highest similarity score.
5. Get the aspect with the highest similarity.
6. The output is a list of aspects extracted from the text.
7. The flow ends.

The results of the sentence embedding process are shown in Table 4.

The term extraction aspect table calculates the cosine similarity of each sentence. The structure of words that are potentially included in the "service" and "general" aspects are as follows:

1. "complimentary noodles", related to the "service" aspect because it concerns the service or action expected from the restaurant.
2. "repeated requests for sugar", related to the "service" aspect because it highlights the interaction of customers with staff or waiters.
3. "dishes", More likely to fall into the "general" aspect because it refers to products or services provided by restaurants in general.

According to the research, the category aspects of each term aspect are grouped more specifically; the grouping is shown in Table 5.

3.3 Aspect sentiment classification (ASC)

Using transformers such as BERT or BIGBIRD, we can parse texts like "***Judging from previous posts this used to be a good place, but not any longer***" more thoroughly. These models can understand context and nuances in text better than traditional approaches.

Table 5. List Aspect Category.

ID	Aspect Category	Entity
1	FOOD#QUALITY	FOOD
2	RESTAURANT#GENERAL	RESTAURANT
3	SERVICE#GENERAL	SERVICE
4	AMBIENCE#GENERAL	AMBIENCE
5	RESTAURANT#MISCELLANEOUS	RESTAURANT
6	FOOD#STYLE_OPTIONS	FOOD
7	RESTAURANT#PRICES	RESTAURANT
8	DRINKS#QUALITY	DRINKS
9	FOOD#PRICES	FOOD
10	LOCATION#GENERAL	LOCATION
11	DRINKS# STYLE_OPTIONS	DRINKS
12	DRINKS#PRICES	DRINKS

Table 6. Score sentiment

Clean Text	Sentiment	Sentiment Propabilitas
Judging from previous posts this used to be a good place, but not any longer.	Negative	[1.4543, 0.0255, -0.7750]
how fun was dry; pork shu mai was more than usually greasy and had to share a table with loud and rude family.	Negative	[3.4650, -0.7887, -1.7910]
The ambience is pretty and nice for conversation, so a casual lunch here would probably be best.	Positive	[-1.9697, -0.9077, 2.9084]

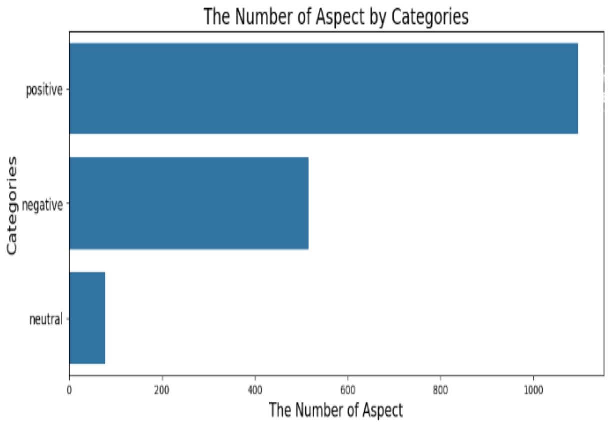

Figure. 4 Number aspect sentiment

This sentence looks positive because it describes the place as "**a good place**" in the past, However, using the transformer approach, the model can recognize the change in sentiment that occurs, namely from positive to negative with the phrase "**but not any longer**". Thus, transformer algorithms can help us understand the change in sentiment hidden in the text, providing a deeper understanding of the evaluation and views contained in the sentence.

Untuk sebaran data dengan label Aspect Sentiment di tunjukan pada Fig. 4.

The next process combines Sentence Transfomers BERT (SBERT), Tagger Part-of-speech, and the Sparse attention mechanisms (BIGBIRD) method for training to obtain better accuracy when tested using the validation dataset.

Received: May 26, 2024. Revised: August 17, 2024.

4. Result and analysis

This study uses two testing models: sentence meaning testing and ATE and ASC task testing, using a combination of Sentence Transfomers SBERT, Tagger part of speech, and the sparse attention mechanism BIGBIRD.

4.1 Testing the meaning of sentences and testing complexity

ABSA problem preprocessing stages such as punctuation removal, lowercase changes, and conjunction removal. These stages have the potential to eliminate the meaning of sentences. This test uses cosine similarity testing between sentence embedding SBERT, word embedding BERT and

Table 8. Similarity scores for each aspect

Sentence Embeding	Aspect term extraction	Similarity scores for each aspect
['judging', 'from', 'previous', 'posts', 'this', 'used', 'to', 'be', 'a', 'good', 'place', 'but', 'not', 'any', 'longer']	Location	FOOD#QUALITY: 0.03394868224859238 RESTAURANT#GENERAL: 0.7744936943054199 SERVICE#GENERAL: 0.6648441553115845 AMBIENCE#GENERAL: 0.2551043629646301 RESTAURANT#MISCELLANEOUS: 0.7593094110488892 FOOD#STYLE_OPTIONS: -0.01818837597966194 RESTAURANT#PRICES: 0.4760243892669678 DRINKS#QUALITY: -0.13794755935668945 FOOD#PRICES: 0.09281834959983826 LOCATION#GENERAL: 0.7852946519851685 DRINKS#STYLE_OPTIONS: -0.1564464271068573 DRINKS#PRICES: -0.07085591554641724
['we', 'there', 'were', 'four', 'of', 'us', 'arrived', 'at', 'noon', 'the', 'place', 'was', 'empty', 'and', 'the', 'staff', 'acted', 'like', 'we', 'were', 'imposing', 'on', 'them', 'and', 'they', 'were', 'very', 'rude']	Service	FOOD#QUALITY: 0.0324741005897522 RESTAURANT#GENERAL: 0.3654351234436035 SERVICE#GENERAL: 0.8794914484024048 AMBIENCE#GENERAL: 0.16239486634731293 RESTAURANT#MISCELLANEOUS: 0.2884083092212677 FOOD#STYLE_OPTIONS: 0.00887487456202507 RESTAURANT#PRICES: 0.1628820151090622 DRINKS#QUALITY: -0.025943582877516747 FOOD#PRICES: 0.017050936818122864 LOCATION#GENERAL: 0.36414429545402527 DRINKS#STYLE_OPTIONS: -0.05704132467508316 DRINKS#PRICES: 0.008699771016836166
['they', 'never', 'brought', 'us', 'compliment', '##ary', 'noodles', 'ignored', 'repeated', 'requests', 'for', 'sugar', 'and', 'threw', 'our', 'dishes', 'on', 'the', 'table']	noodles, requests, sugar, dishes	FOOD#QUALITY: 0.7418381571769714 RESTAURANT#GENERAL: -0.017319289967417717 SERVICE#GENERAL: 0.1242925375699997 AMBIENCE#GENERAL: -0.1709030270576477 RESTAURANT#MISCELLANEOUS: -0.08848093450069427 FOOD#STYLE_OPTIONS: 0.8362204432487488 RESTAURANT#PRICES: -0.2351987361907959 DRINKS#QUALITY: -0.05219016969203949 FOOD#PRICES: 0.7052836418151855 LOCATION#GENERAL: -0.05418545752763748 DRINKS#STYLE_OPTIONS: -0.16691933572292328 DRINKS#PRICES: -0.2543781101703644

Table 8. Evaluation of Aspect Term Extraction.

Aspect Term Category	Precision	Recall	F1-score
FOOD#QUALITY	0.94	0.97	0.95
RESTAURANT#GENERAL	0.93	0.93	0.93
SERVICE#GENERAL	1.00	0.95	0.97
AMBIENCE#GENERAL	0.86	0.96	0.91
RESTAURANT#MISCELLANEOUS	0.91	0.73	0.81
FOOD#STYLE_OPTIONS	0.71	0.71	0.71
RESTAURANT#PRICES	0.87	1.00	0.93
DRINKS#QUALITY	0.71	0.83	0.76
FOOD#PRICES	0.66	0.66	0.66
LOCATION#GENERAL	0.66	0.66	0.66
DRINKS#STYLE_OPTIONS	1.00	0.50	0.66
DRINKS#PRICES	0.00	0.00	0.00

LSTM. The result prove that SBERT can outperform BERT and LSTM in testing sentence meanings. SBERT has an accuracy of **0.971**, BERT has an accuracy of **0.963** and LSTM **0.876** for recognizing sentence meaning using the hotel review dataset [9, 10, 13-15].

For time complexity and memory usage testing, BIGBIRD is faster and uses less memory than BERT.

The time complexity of BERT is **10.85 seconds**, BIGIRD is **8.02 seconds**. The space complexity of BERT requires **604.87 Mb**, BIGBIRD **602.84 Mb**. From these tests, BIGIRBIRD can be applied to ABSA tasks, especially ATE and ASC.

4.2 Aspect term extraction result

The term extraction results can be continued to the category aspect process by calculating the cosine similarity of the sentence embedding using SBERT compared to Location, which will produce a category aspect score. Table 7 is the result of the category aspect and cosine similarity score.

The results of the evaluation of each aspect term category are shown in Table 8.

For the evaluation of the aspect term, each precision, recall, f1-score, and overall accuracy are produced.

In analyzing various aspects of the restaurant, we obtained interesting results. The food aspect (FOOD) showed very good performance, with a precision level of 90%, a recall level of 95%, and an F1-score of 92%. This indicates that our model is able to identify and evaluate the food aspect well in restaurant reviews.

Meanwhile, the service aspect (SERVICE) also received high ratings, with a precision of 94% a recall of 90%, and an F1-score of 92%. However, there are other aspects, such as atmosphere (AMBIENCE), which showed slightly lower results with a precision of 75%, a recall of 84%, and an F1-score of 79%. Furthermore, the beverage aspect (DRINKS) showed very high precision, reaching 100%, but its recall was low at 72%, and an F1-score of 84%.

This study also outperforms the accuracy of the LDA+LSTM [13], PLSA+LSTM [14], BERT+LDA [9], and Attention-based Sentence + BERT [10] studies when tested using the same dataset in the study.

Table 9. Evaluation of Aspect Category.

Aspect Entity	Precision	Recall	F1-score
FOOD	0.90	0.95	0.92
RESTAURANT	0.88	0.86	0.87
SERVICE	0.94	0.90	0.92
AMBIENCE	0.75	0.84	0.79
DRINKS	1.00	0.72	0.84
LOCATION	0.00	0.00	0.00

Table 10. Comparison of Aspect Term Extraction.

Aspect Sentiment	Accuracy
LDA-LSTM [13]	0.93
PLSA-LSTM [14]	0.94
BERT-LDA [9]	0.97
Attention Based Sentence-BERT [10]	0.98
Propshed Method (SBERT+Tagger Part-of-speech+Sparse Attention Mechanism)	**0.99**

Table 11. Comparison of Aspect Sentiment Classification.

Aspect Sentiment Classification	Accuracy
Evaluation of weakly-supervised methods for aspect extraction [20]	0.60
Rhetorical Structure Theory [21]	0.79
Combination of Recursive and RNNs [22]	0.80
Deep Learning for Multilingual Aspect-based Sentiment Analysis [23]	0.82
Recursive neural conditional random fields for ABSA [28]	0.84
ALDONA [24]	0.86
Propshed Method (SBERT+Sparse Attention Mechanism)	**0.89**

4.3 Aspect sentiment classification result

The process for sentiment polarity, using a combination of Sentence Transfomers SBERT and Sparse attention mechanism BIGBIRD with a model that has been trained in advance, is proven to recognize a sentence more. This study conducted a special ASC test using the 2016 semeval dataset, and the results were higher when compared to several previous studies.

ASC research with the proposed method as shown in Table 11 shows better accuracy than previous research, the accuracy results are **0.893**, precision **0.883**, recall **0.893**, and F1-score **0.882**.

5. Conclusion

This study focuses on two important aspects of aspect-based sentiment analysis: Aspect Term Extraction (ATE) and Aspect Sentiment Classification (ASC). The first stage of this study is to compare the use of word tokenization with

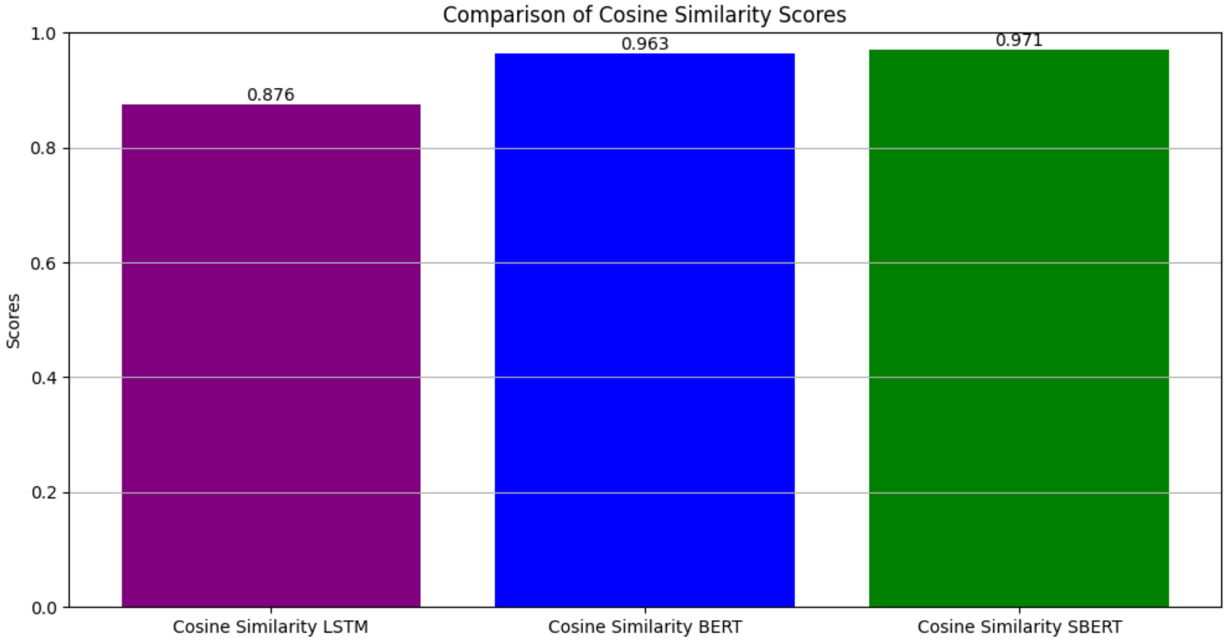

Figure. 5 Comparison Cosine Similarity LSTM, BERT, SBERT

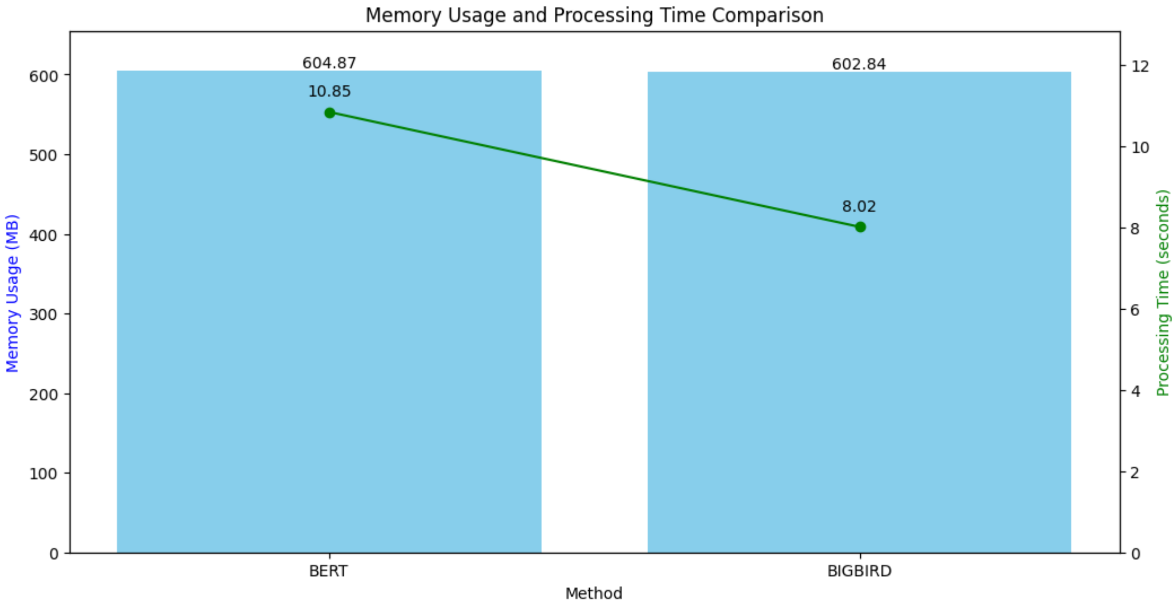

Figure. 6 Time Complexity and Space Complexity between (BERT dan BIGBIRD)

sentence tokenization. Word tokenization uses BERT, while this study uses a combination of BERT, Tagger Part-of-Speech (POS), and Cosine Similarity (SBERT) calculations. To test whether SBERT sentence tokenization is better than word tokenization using BERT. Testing the meaning of sentences is shown in Fig. 5. The test uses the same dataset as the studies [9, 10, 13-15].

The testing method used is cosine similarity to produce a semantic similarity score. The test results are LSTM **0.876**, BERT **0.963** and SBERT **0.971** SBERT is superior in recognizing sentence meaning so that it can be a reference in the use of further methods. The second test is the time complexity test and memory usage test on the BERT and BIGBIRD classification methods. The test results are shown in Fig. 6. BERT **10.85 seconds**, BIGBIRD **8.02 seconds**. BERT space complexity requires **604.87 Mb**, BIGBIRD **602.84 Mb**. BIGBIRD is proven to be more effective than BERT.

After getting better sentence embedding tokenization than word embedding, continued with ATE testing, By combining Sentence BERT (SBERT), Tagger Part of speech (POS), and the BIGBIRD sparse attention mechanism method to be tested with research [13,14,9,10] the results are shown in Table 10 the combination of models proposed by this study outperforms previous research, getting an accuracy of **0.99**.

The second test uses a different dataset from the first test while also trying to determine whether the proposed method can work well on different domains and different tasks. The second test uses the dataset [20-24, 28] with an accuracy score of **0.893** as shown in Table 11. It turns out that the combination of Sentence BERT (SBERT), Tagger POS, and Sparse attention mechanism BIGBIRD can be used on more specific ABSA tasks, namely ATE and ASC. However, there are still limitations in this study, such as the size of the dataset used, the type of dataset domain that is more varied, multi-language, not only English, and the possibility of overfitting in the trained model. Future research is expected to explore various types of datasets, other model approaches, the use of imbalance dataset techniques, and regularization techniques to overcome overfitting and experiment with larger and more diverse datasets.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, experiment, writing draft preparation done by 1st and 2nd authors. The validation, analysis, writing, review and editing have been done by 3nd and 4nd authors.

Acknowledgments

This work funded by the Ministry of Education, Culture, Research, and Technology Indonesia under Penelitian Disertasi Doktor 2024 [Grant Research Number : 108/E5/PG.02.00.PL/2024]

References

[1] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges”, *IEEE Trans Knowl Data Eng*, Vol. 35, No. 11, pp. 11019-11038, 2023, doi: 10.1109/TKDE.2022.3230975.

[2] S. Bhatia, M. Sharma, and K. K. Bhatia, “Sentiment Analysis and Mining of Opinions”, *Studies in Big Data*, Vol. 30, No. May, pp. 503-523, 2018, doi: 10.1007/978-3-319-60435-0_20.

[3] T. A. Rana and Y. N. Cheah, “Sequential patterns rule-based approach for opinion target extraction from customer reviews”, *J Inf Sci*, Vol. 45, No. 5, pp. 643-655, 2019, doi: 10.1177/0165551518808195.

[4] M. Pontiki, “SemEval-2016 task 5: Aspect based sentiment analysis”, *Association for Computational Linguistics*, pp. 19-30, 2016, doi: 10.18653/v1/s16-1002.

[5] Ł. Augustyniak, T. Kajdanowicz, and P. Kazienko, “Comprehensive analysis of aspect term extraction methods using various text embeddings”, *Comput Speech Lang*, Vol. 69, 2021, doi: 10.1016/j.csl.2021.101217.

[6] H. Xu, B. Liu, L. Shu, and P. S. Yu, “Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction”, In: *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (Short Papers)*, Vol. Proceeding, pp. 592-598, 2018, doi: 10.18653/v1/P18-2094.

[7] M. Orbach, O. Toledo-Ronen, A. Spector, R. Aharonov, Y. Katz, and N. Slonim, “YASO: A Targeted Sentiment Analysis Evaluation Dataset for Open-Domain Reviews”, In: *Proc. of EMNLP 2021-2021 Conference on Empirical Methods in Natural Language Processing*, pp. 9154-9173, 2021, doi: 10.18653/v1/2021.emnlp-main.721.

[8] A. Karimi, L. Rossi, and A. Prati, “Improving BERT Performance for Aspect-Based Sentiment Analysis”, In: *Proc. of the 4th International Conference on Natural Language and Speech Processing*, pp. 39-46, 2021.

[9] P. N. Andono, Sunardi, R. A. Nugroho, and B. Harjo, "Aspect-Based Sentiment Analysis for Hotel Review Using LDA, Semantic Similarity, and BERT", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 5, pp. 232-243, 2022, doi: 10.22266/ijies2022.1031.21.

[10] B. Harjo, Muljono, and R. Abdullah, "Attention-based Sentence Extraction for Aspect-based Sentiment Analysis with Implicit Aspect Cases in Hotel Review Using Machine Learning Algorithm, Semantic Similarity, and BERT", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 3, pp. 189-200, 2023, doi: 10.22266/ijies2023.0630.15.

[11] Suhariyanto, R. Abdullah, R. Sarno, and C. Fatichah, "Aspect-Based Sentiment Analysis for Sentence Types with Implicit Aspect and Explicit Opinion in Restaurant Review Using Grammatical Rules, Hybrid Approach, and SentiCircle", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, pp. 177-187, 2021, doi: 10.22266/ijies2021.1031.17.

[12] R. Abdullah, Suhariyanto, and R. Sarno, "Aspect Based Sentiment Analysis for Explicit and Implicit Aspects in Restaurant Review using Grammatical Rules, Hybrid Approach, and SentiCircle", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, pp. 294-305, 2021, doi: 10.22266/ijies2021.1031.27.

[13] R. A. Priyantina, "Sentiment analysis of hotel reviews using Latent Dirichlet Allocation, semantic similarity and LSTM", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 4, pp. 142-155, 2019, doi: 10.22266/ijies2019.0831.14.

[14] D. Khotimah and R. Sarno, "Sentiment Analysis of Hotel Aspect Using Probabilistic Latent Semantic Analysis, Word Embedding and LSTM", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 4, pp. 275-290, Aug. 2019, doi: 10.22266/ijies2019.0831.26.

[15] M. D. Pratama, R. Sarno, and R. Abdullah, "Sentiment Analysis User Regarding Hotel Reviews by Aspect Based Using Latent Dirichlet Allocation, Semantic Similarity, and Support Vector Machine Method", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 3, pp. 514-524, 2022, doi: 10.22266/ijies2022.0630.43.

[16] M. Zaheer *et al.*, "Big bird: Transformers for longer sequences", *Adv Neural Inf Process Syst*, Vol. 2020-Decem, No. NeurIPS, 2020.

[17] D. Marutho, Muljono, S. Rustad, and Purwanto, "Optimizing aspect-based sentiment analysis using sentence embedding transformer, bayesian search clustering, and sparse attention mechanism", *Journal of Open Innovation: Technology, Market, and Complexity*, Vol. 10, No. 1, p. 100211, 2024, doi: 10.1016/j.joitmc.2024.100211.

[18] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier", *J Phys Conf Ser*, Vol. 1192, No. 1, 2019, doi: 10.1088/1742-6596/1192/1/012024.

[19] *et al* Divya Khyani, Siddhartha B S, "An Interpretation of Lemmatization and Stemming in Natural Language An Interpretation of Lemmatization and Stemming in Natural Language Processing", *Journal of University of Shanghai for Science and Technology*, Vol. 22, No. 10, p. 350, 2021.

[20] M. Ettaleb, A. Barhoumi, N. Camelin, and N. Dugué, "Evaluation of weakly-supervised methods for aspect extraction", *Procedia Comput Sci*, Vol. 207, No. Kes, pp. 2688-2697, 2022, doi: 10.1016/j.procs.2022.09.327.

[21] K. Schouten and F. Frasincar, "COMMIT at SemEval-2016 task 5: Sentiment analysis with rhetorical structure theory", *SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings*, pp. 356-360, 2016, doi: 10.18653/v1/s16-1057.

[22] C. R. Aydln, "Combination of recursive and recurrent neural networks for aspect-based sentiment analysis using inter-aspect relations", *IEEE Access*, Vol. 8, pp. 77820-77832, 2020, doi: 10.1109/ACCESS.2020.2990306.

[23] S. Ruder, P. Ghaffari, and J. G. Breslin, "INSIGHT-1 at SemEval-2016 Task 5: Deep learning for multilingual aspect-based sentiment analysis", In: *Proc. of SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings*, pp. 330-336, 2016, doi: 10.18653/v1/s16-1053.

[24] D. Meškelė, "ALDONAr: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model", *Inf Process Manag*, Vol. 57, No. 3, 2020, doi: 10.1016/j.ipm.2020.102211.

[25] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language

Received: May 26, 2024. Revised: August 17, 2024.

understanding", In: *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, Vol. 1, No. Mlm, pp. 4171-4186, 2019.

[26] Z. Gao, "Target-dependent sentiment classification with BERT", *IEEE Access*, Vol. 7, pp. 154290-154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[27] and I. A. Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Suresh Manandhar, "SemEval-2016 Task 5", In: *Proc. of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver*, Colorado. [Online]. Available: https://alt.qcri.org/semeval2016/task5/

[28] W. Wang, "Recursive neural conditional random fields for aspect-based sentiment analysis", *Association for Computational Linguistics*, pp. 616-626 2016, doi: 10.18653/v1/d16-1059.