JN-NLP at SIGHAN-2024 dimABSA Task: Extraction of Sentiment Intensity Quadruples Based on Paraphrase Generation

Yunfan Jiang and **Tianci Liu** and **Hengyang Lu**
School of Artificial Intelligence and Computer Science, Jiangnan University, China
{1033200623,liutianci}@stu.jiangnan.edu.cn
luhengyang@jiangnan.edu.cn

Abstract

Aspect-based sentiment analysis (ABSA) is a fine-grained sentiment analysis task, which aims to extract multiple specific sentiment elements from text. The current aspect-based sentiment analysis task mainly involves four basic elements: aspect term, aspect category, opinion term, and sentiment polarity. With the development of ABSA, methods for predicting the four sentiment elements are gradually increasing. However, traditional ABSA usually only distinguishes between "positive", "negative", or "neutral" attitudes when judging sentiment polarity, and this simplified classification method makes it difficult to highlight the sentiment intensity of different reviews. SIGHAN 2024 provides a more challenging evaluation task, the Chinese dimensional ABSA shared task (dimABSA), which replaces the traditional sentiment polarity judgment task with a dataset in a multidimensional space with continuous sentiment intensity scores, including valence and arousal. Continuous sentiment intensity scores can obtain more detailed emotional information. In this task, we propose a new paraphrase generation paradigm that uses generative questioning in an end-to-end manner to predict sentiment intensity quadruples, which can fully utilize semantic information and reduce propagation errors in the pipeline approach.

1 Introduction

Traditional Aspect-based sentiment analysis (ABSA) can extract four specific emotional elements from the text: 1) *Aspect term*, which is a specific aspect in the sentence, generally a word or phrase expressed in the text, and may not exist; 2) *Aspect category*, the category involved by the aspect term, usually a predefined set of categories; 3) *Opinion term*, the expression of a specific emotional view on an aspect; 4) *Sentiment polarity*, the emotional tendency towards a certain aspect. For example, for the review "*The pizza at this restaurant is delicious, but the service is terrible.*", the ABSA task can extract two emotional quadruples: (*pizza*, *food type*, *delicious*, *positive*) and (*service*, *service attitude*, *terrible*, *negative*).

The Chinese dimensional ABSA shared task (dimABSA) (Lee et al., 2024) dataset is a collection of comments extracted by organizers from online catering industry social media platforms. After removing HTML tags and multimedia tags, the text was split into multiple sentences. A selection of these sentences was then manually annotated with aspect term, aspect category, opinion term, and sentiment intensity. For sentiment intensity, the organizers used the valence and arousal provided by the "Chinese EmoBank" (Lee et al., 2022;Yu et al., 2016) to represent emotional states as continuous numerical values in a multidimensional space. Using valence-arousal as sentiment intensity, this method provides more detailed emotional information. The "Chinese EmoBank" is a manually annotated Chinese emotional dictionary. In it, valence describes the positivity or negativity of emotions, ranging continuously from negative (such as *sadness*, *anger*) to positive (such as *happiness*, *excitement*). Valence is often seen as the "pleasantness" of an emotional experience and is a standard for assessing the quality of emotional experiences. Arousal, on the other hand, refers to the level of excitement of an emotion. It reflects the intensity of an individual's physiological and psychological response to an emotional stimulus. Arousal is also a continuous range, from very low (such as *boredom*, *tiredness*) to very high (such as *surprise*, *panic*). In the "Chinese EmoBank", both valence and arousal are measured on a scale from 1 to 9. In the sentiment intensity of the dimABSA dataset, valence and arousal are separated by a '#'.

The dimABSA task provides three subtasks: Subtask 1: For a given sentence and its aspect term, predict the sentiment intensity of the aspect term

Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), pages 121–126

in the comment; Subtask 2: For a given sentence, extract the aspect term and opinion term from the comment and predict the sentiment intensity; Subtask 3: For a given sentence, extract the aspect term, aspect category, and opinion term from the comment and predict the sentiment intensity. In this task, we have implemented Subtask 3, where the system could extract all sentiment quadruples (aspect, category, opinion, intensity). For example, for the comment "*This bowl of ramen is super invincibly thunderously bad.*", the final extraction of the sentiment intensity quadruple would be: (*ramen*, *food#quality*, *super invincibly thunderously bad*, *2.00#7.88*).

This work processes the dataset provided by the dimABSA task and proposes a new paraphrase generation paradigm. It replaces the traditional sentiment polarity judgment task in ABSA with the judgment of sentiment intensity: valence#arousal, using the T5 pre-trained model (Raffel et al., 2020) that unifies natural language processing tasks into text-to-text tasks. Through fine-tuning training, it accomplishes the task of extracting sentiment intensity quadruples. Experiments demonstrate that our newly proposed paraphrase generation paradigm achieves good performance in predicting sentiment intensity quadruples. Our contributions are summarized as follows: 1) We propose transforming the dimABSA task into a paraphrase generation problem and introduce a new paraphrase generation paradigm, allowing us to fully utilize semantic information while predicting sentiment intensity quadruple in one shot; 2) Our model has been experimentally validated to perform excellently in extracting aspect term, aspect category, and opinion term.

2 Related Work

With the emergence of pre-trained language models like BERT, research in ABSA has made significant progress. Sun et al. (2019) proposed a method based on pre-trained language models, utilizing models like BERT to extract the subjects and aspects from online comments, and employing multi-task learning to determine their sentiment polarity. This was the first method for ABSA based on pre-trained language models and is representative of methods based on transfer learning. Li et al. (2019) were the first to apply BERT to end-to-end ABSA tasks, achieving the best results at the time with a simple linear classifier. Subsequently, composite ABSA tasks began to develop. Researchers proposed various end-to-end models for extracting multiple sentiment elements, capable of handling multiple subtasks in sentiment analysis tasks, such as Aspect Term Extraction (ATE) and Aspect Sentiment Classification (ASC). This integrated approach reduced error propagation and improved overall performance. Liu et al. (2021) adopted the Seq2Seq modeling paradigm to extract aspect category and sentiment polarity, based on pre-trained generative models, using natural language sentences to represent the desired output for Aspect Category Sentiment Analysis (ACSA) tasks. Peng et al. (2020) proposed a two-stage pipeline method for extracting aspect term, opinion term, and sentiment polarity to address the Aspect Sentiment Triplet Extraction (ASTE) task; Wan et al. (2020) introduced the Target-aspect-sentiment joint detection task for aspect-based sentiment analysis (TASD), aimed at simultaneously predicting aspect category, aspect term, and sentiment polarity.

The Aspect Sentiment Quad Prediction (ASQP) task aims to extract the four sentiment elements of a specific sentence at once, revealing a more comprehensive and complete aspect-level sentiment structure. Zhang et al.(2021) proposed a Paraphrase Generation paradigm to solve the ASQP task in English. This approach generates natural language sentences from sentiment quadruples using pre-established templates, making the generated natural language sequence the target sequence, which forms a mapping relationship with the original review sentence. Zhang et al. transformed the original quad element prediction task into a text generation problem, which was then solved using a sequence-to-sequence (Seq2Seq) approach. Compared to the pipeline method, the Seq2Seq approach can reduce the cumulative propagation error caused by accuracy errors at each step in the pipeline, and since the subtasks of ASQP are usually expressed as token-level or sequence-level classification problems, the Seq2Seq approach can make full use of semantic information.

3 Methodology

3.1 Problem Statement

Subtask 3 of the dimABSA task involves extracting a sentiment intensity quadruple (a, c, o, i) from a given sentence, which corresponds to aspect term, aspect category, opinion term, and sentiment intensity, respectively. The aspect term in the original comment sentence may not exist, and when it is absent, 'NULL' is used as the aspect term. The aspect category in the original comment are predefined, and the dimABSA task divides comments on different aspects of the catering industry into twelve categories, each with one entity and one attribute corresponding to the aspect term. The aspect term can be directly extracted from the comment. Sentiment intensity is divided into valence and arousal.

The dimABSA task dataset provides 6050 training entries, 2000 test entries, and 100 validation entries.

3.2 ASQP Task

To highlight the main content of the sentiment quadruple, Zhang et al.'s paraphrase generation task linearizes the sentiment quadruple $Q = (c, a, o, p)$ into a natural sentence as follows:

$$P_c(c)\, is\, P_p(p)\, because\, P_a(a)\, is\, P_o(o)$$

Herein,$P_z(\cdot)$ belongs to the mapping function of $z \in \{c, a, o, p\}$, which maps the sentiment element z from its original format into natural language form. In the sentiment quadruple, c and o are already in natural language form. As for the sentiment polarity, its mapping is as follows:

$$P_p(p) = \begin{cases} great & if\, p = positive \\ ok & if\, p = neutral \\ bad & if\, p = negative \end{cases} \quad (1)$$

Aspect term may not exist in the original sentence, in which case they are considered as an implicit aspect term 'NULL'; otherwise, they are in natural language form. Their mapping method is as follows:

$$P_a(a) = \begin{cases} it & if\, a = NULL \\ a & otherwise \end{cases} \quad (2)$$

When Zhang et al. handle Seq2Seq learning, for a given sentence x, the encoder first converts it into a contextualized encoded sequence e. Then, the decoder simulates the conditional probability distribution of the target sentence y given the encoded input representation: $P_\theta(y|e)$, which is parameterized by θ.

At the i-th time step, the output y_i of the decoder is based on the encoded input sequence e and the previous output $y_{<i}$, where $y_{<i} : y_i = f_{dec}(e, y_{<i})$, and $f_{dec}(\cdot)$ represents the computed value of the decoder. To obtain the probability distribution of the next token, the following softmax function is applied:

$$P_\theta(y_{i+1}|e, y_{<i+1}) = softmax(W^T y_i) \quad (3)$$

Here, W maps the predicted value y_i to a logit vector, which is defined as the logarithmic odds ratio of an event occurring versus not occurring. It can be used to calculate the probability distribution over the entire vocabulary set. The formula is as follows:

$$logit(P) = log(\frac{P}{1-P}) \quad (4)$$

Training with the T5 pre-trained model can achieve the initialization of pre-trained parameter weights θ, and further fine-tune the input-target pairs, thus maximizing the probability distribution $P_\theta(y|e)$:

$$\underset{\theta}{max}\, logP_\theta(y|e) = \sum_{i=1}^{n} logP_\theta(y_i|e, y_{<i}) \quad (5)$$

where n is the length of the sequence target y.

3.3 DimABSA as Paraphrase Generation

Regarding the 'Sentence', 'Aspect', 'Category', 'Opinion', and 'Intensity' aspects in the dimABSA task dataset, the data format is processed into the following format:

$$Sentence[[A_1, C_1, O_1, I_1], ...[A_n, C_n, O_n, I_n]]$$

We propose a new paraphrase generation paradigm to handle the task of Chinese sentiment intensity quadruple extraction, for a given sentence tuple pair (x,Q), with the goal of generating a target pair of sentences in Chinese natural language (x,y), the sentiment intensity quadruple is linearized into a natural sentence as follows:

$$P_C(C)\, valence\, is\, P_I(I-v)\, arousal$$
$$is\, P_I(I-a)\, because\, P_A(A)\, is\, P_O(O)$$

Wherein,$P_z(\cdot)$ is the mapping function $z \in \{C, A, O, I\}$, which transforms sentiment elements from their original format into natural language form. In the sentiment intensity quadruple, the aspect category (C) and opinion term (O) are already in natural language form. The aspect term (A) may not exist in the original sentence, in which case it is considered an implicit aspect term

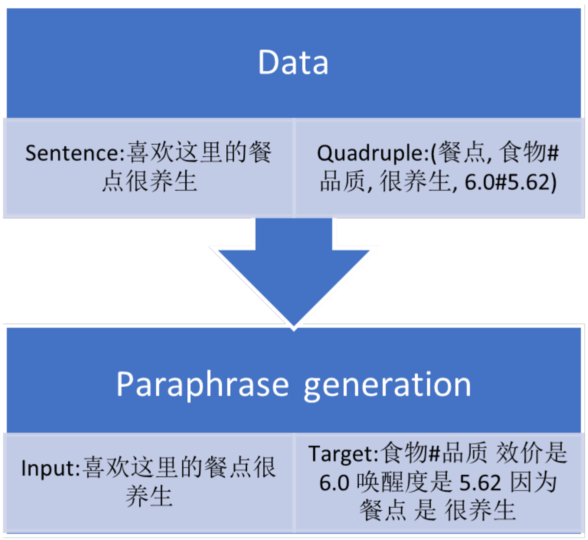

Figure 1: An example of sentence tuple pairs generating sentence target pairs.

'NULL', otherwise, it is directly regarded as natural language form. Its mapping method is as follows:

$$P_A(A) = \begin{cases} it & if\, a = NULL \\ a & otherwise \end{cases} \quad (6)$$

The sentiment intensity (I) will be directly retained as a scoring standard in the natural language generated after paraphrase generation, representing the valence and arousal ratings for the given aspect term. The final target sentence y generated can form a mapping relationship with the original comment sentence x, which is then directly fine-tuned using the T5 pre-trained language model. When there are multiple sentiment intensity quadruples Q in a sentence x, the separator [SSEP] is used to divide the multiple target sentences generated. Figure 1 shows an example of paraphrase generation.

4 Experiment

4.1 Experiment Details

The mt5-base (Xue et al., 2021) is a model proposed by Google, pre-trained on the mC4 corpus, and includes 101 languages, including Chinese. In this work, the t5-base-chinese pre-trained model is selected as the task model. The t5-base-chinese is based on mt5-base, retaining only Chinese and English for pre-training. Fine-tune training on T5-base-Chinese. Both training and evaluation batch sizes are set to 16; gradient_accumulate_steps is set to 1, the learning rate is set to 3e-4, and the number of training rounds is set to 10.

4.2 Main Results

Organizers use accuracy, recall, and F1-score to evaluate the model. The higher these three metrics, the better the model's performance. A quadruple is regarded as correct if and only if the four elements and their combination match those in the gold quadruple. Table 1 shows the scores of the three metrics: accuracy, recall, and F1-score, for valence and arousal in this work.

A total of 7 teams submitted, and the published F1-scores for valence, arousal, and valence-arousal are as shown in Table 2. Our work is ranked fifth.

At the same time, this work also trained a T5 pre-trained model without using the paraphrase generation paradigm, to comparatively evaluate the model's ability to extract aspect term, aspect category, and opinion term in sentiment triplets. The test model, which does not use paraphrase generation, directly maps the golden triplets containing aspect term, aspect category, and opinion term as the target sequence, forming a mapping with the original comment sentence. The final test model outputs predicted triplets for the input comments. After manually removing some problematic test data, the performance of this test model and the task model on 900 test data in terms of accuracy, recall rate, and F1-score for the aforementioned three types of sentiment elements is shown in Table 3.

4.3 Error Analysis

When the model processes some more complex natural language sentences, it outputs some problematic target sentences: 1) 'The watermelon and strawberries are very fresh and delicious.', this sentence in chinese contains two aspect terms and two opinion terms. Both opinion terms are expressions of sentiment for the two aspect terms, and there is no conjunction between the two opinion terms. The model has difficulty correctly matching aspect terms with opinion terms for comments that have multiple aspect terms and opinion terms without direct conjunctions, resulting in the repeated output of the same target sentence: 'Food#quality valence is 6.75 arousal is 6.25 because watermelon is delicious'. When a conjunction is added between the two opinion terms, the model's output is normal; 2) Due to the diversity of Chinese language forms, the model struggles to process some idioms or longer comments. In the output of target sentences, it will manifest as the target sentence not conforming to the rules of paraphrase generation paradigms, such

	Precision	Recall	F1-score
Valence	0.484	0.480	0.482
Arousal	0.441	0.437	0.439
V-A	0.333	0.330	0.331

Table 1: The scores of the model's valence, arousal, and valence-arousal on accuracy, recall, and F1-score.

Team	Valence-F1	Arousal-F1	V-A-F1
HITSZ-HLT	**0.567**	**0.526**	**0.417**
CCIIPLab	0.555	0.507	0.389
ZZU-NLP	0.522	0.489	0.376
SUDA-NLP	0.487	0.444	0.336
JN-NLP (ours)	0.482	0.439	0.331
BIT-NLP	0.470	0.434	0.329
USTC-NLP	0.438	0.437	0.312

Table 2: The F1-scores of the participating teams in valence, arousal, and valence-arousal.

as the absence of a certain sentiment element in the target sentence, or the appearance of multiple setting words from paraphrase generation paradigms in one target sentence, such as 'valence is', 'arousal is', etc. The model still has limitations in the above examples.

5 Conclusions

The goal of dimABSA subtask 3 is to extract sentiment intensity quadruples from online review sentences in the catering industry, including aspect term, aspect category, opinion term, and the valence-arousal representing sentiment intensity. This work proposes a new paraphrase generation paradigm, utilizing the dataset provided by the dimABSA task, and ultimately achieves a model based on the T5 pre-trained model fine-tuned for training. This model uses the new paraphrase generation paradigm to facilitate Seq2Seq learning, transforming sentiment intensity quadruples into natural language target sentences, forming a mapping relationship with the review sentences. The model can generate an output sentence according to the paraphrasing rules for the input sentence, and the sentiment intensity quadruples can be obtained by processing the output sentence. By comparing the performance of this work's model with the test model in extracting triples of aspect term, aspect category, and opinion term, it is evident that this work's model performs better.

Model	precision	recall	F1-score
test model	0.434	0.395	0.414
JN-NLP (ours)	**0.462**	**0.470**	**0.466**

Table 3: Comparison of the test model and our model in terms of accuracy, recall, and F1-score on predicting sentiment triplets.

Limitations

After training the model with the paraphrase generation paradigm we proposed, it can complete the task of extracting the sentiment intensity quadruples. However, the model still has limitations in predicting sentiment intensity. When a comment contains multiple aspect terms, the model may predict the same score for the sentiment intensity of multiple aspect terms. Moreover, due to the complexity of the Chinese language, the model may generate incorrect target sentences for some Chinese comments.

Acknowledgments

This research was funded by the China Postdoctoral Science Foundation (Grant No. 2022M711360) and the Laboratory for Advanced Computing and Intelligence Engineering.

References

[1] Lung-Hao Lee, Jian-Hong Li and Liang-Chih Yu, "Chinese EmoBank: Building Valence-Arousal Resources for Dimensional Sentiment Analysis," *ACM Trans. Asian and Low-Resource Language Information Processing*, vol. 21, no. 4, article 65, 2022.

[2] Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016. "Building Chinese affective resources in valence-arousal dimensions. In *Proceedings of NAACL/HLT-16*, pages 540-545.

[3] Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. Aspect Sentiment Quad Prediction as Paraphrase Generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[4] Lung-Hao Lee, Liang-Chih Yu, Suge Wang, and Jian Liao. 2024. Overview of the SIGHAN 2024 shared task for Chinese dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.

[5] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou,Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res., 21:140:1–140:67.

[6] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 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, pages 483–498, Online. Association for Computational Linguistics.

[7] Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 380–385, Minneapolis, Minnesota. Association for Computational Linguistics.

[8] X. Li, L. Bing, P. Li, and W. Lam, "A unified modelfor opinion target extraction and target sentiment prediction," in AAAI, 2019, pp. 6714–6721.

[9] Jian Liu, Zhiyang Teng, Leyang Cui, Hanmeng Liu, and Yue Zhang. 2021. Solving Aspect Category Sentiment Analysis as a Text Generation Task. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4406–4416, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[10] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," in AAAI, 2020, pp. 8600–8607.

[11] H. Wan, Y. Yang, J. Du, Y. Liu, K. Qi, and J. Z. Pan, "Target-aspect-sentiment joint detection for aspect-based sentiment analysis," in AAAI, 2020, pp. 9122–9129.