Overview of the SIGHAN 2024 Shared Task for Chinese Dimensional Aspect-Based Sentiment Analysis

Lung-Hao Lee[1], Liang-Chih Yu[2], Suge Wang[3, 4], and Jian Liao[3]

[1]Institute of Artificial Intelligence Innovation, National Yang Ming Chiao Tung University

[2]Department of Information Management, Yuan Ze University

[3]School of Computer and Information Technology, Shanxi University, Taiyuan, China

[4]Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China

lhlee@nycu.edu.tw, lcyu@saturn.yzu.edu.tw, {wsg, liaoj}@sxu.edu.cn

Abstract

This paper describes the SIGHAN-2024 shared task for Chinese dimensional aspect-based sentiment analysis (ABSA), including task description, data preparation, performance metrics, and evaluation results. Compared to representing affective states as several discrete classes (i.e., *sentiment polarity*), the dimensional approach represents affective states as continuous numerical values (called *sentiment intensity*) in the valence-arousal space, providing more fine-grained affective states. Therefore, we organized a dimensional ABSA (shorted dimABSA) shared task, comprising three subtasks: 1) intensity prediction, 2) triplet extraction, and 3) quadruple extraction, receiving a total of 214 submissions from 61 registered participants during evaluation phase. A total of eleven teams provided selected submissions for each subtask and seven teams submitted technical reports for the subtasks. This shared task demonstrates current NLP techniques for dealing with Chinese dimensional ABSA. All data sets with gold standards and evaluation scripts used in this shared task are publicly available for future research.

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) (Pontiki et al., 2014; 2015; 2016) is a critical NLP research topic that aims to identify the aspects of a given entity and analyze the sentiment polarity associated with each aspect. In recent years, considerable research has been devoted to ABSA, which can be categorized into different tasks based on the number of sentiment elements to be extracted. For example, the Aspect Sentiment Triplet Extraction (ASTE) task (Yuan et al., 2023; Chen et al., 2021; Mao et al., 2021; Peng et al., 2020; Wu et al., 2020; Xu et al., 2020; Zhang et al., 2020) extracts three elements in a triplet, including aspect/target term, opinion term and sentiment polarity (e.g., positive, neutral, and negative). Furthermore, the Aspect Sentiment Quadruple Prediction (ASQP) task (Cai et al., 2021; Gao et al., 2022; Mao et al., 2022; Peper and Wang, 2022; Zhang et al., 2021; Zhou et al., 2023) extracts the same three elements plus an additional aspect category to construct a quadruple.

However, compared to representing affective states as several discrete classes (i.e., *sentiment polarity*), the dimensional approach that represents affective states as continuous numerical values (called *sentiment intensity*) in multiple dimensions such as valence-arousal (VA) space (Russel, 1980), providing more fine-grained emotional information (Lee et al., 2022; Deng et al., 2022; 2023; Yu et al., 2016).

Therefore, we organized a Chinese dimensional ABSA shared task (dimABSA) in the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN 2024), providing fine-grained sentiment intensity prediction for each extracted aspect of a restaurant review. We have three subtasks: 1) Intensity Prediction, 2) Triplet Extraction, and 3) Quadruple Extraction. Participants are free to participate in any or all subtasks. Given a sentence with/without aspects, participating systems should be able to extract the

Example	Version	Input & Output
Example 1 (subtask 1)	Traditional	Input: E0001:S001, 檸檬醬也不會太油，塔皮對我而言稍軟。, 檸檬醬#塔皮 Output: E0001:S001 (檸檬醬,5.67#5.50)(塔皮,4.83#5.00)
	Simplified	Input: E0001:S001, 柠檬酱也不会太油，塔皮对我而言稍软。 柠檬酱#塔皮 Output: E0001:S001 (柠檬酱,5.67#5.50)(塔皮,4.83#5.00)
Example 2 (subtask 2)	Traditional	Input: E0002:S002, 不僅餐點美味上菜速度也是飛快耶！！ Output: E0002:S002 (餐點, 美味, 6.63#4.63) (上菜速度, 飛快, 7.25#6.00)
	Simplified	Input: E0002:S002, 不仅餐点美味上菜速度也是飞快耶!! Output: E0002:S002 (餐点, 美味, 6.63#4.63) (上菜速度, 飞快, 7.25#6.00)
Example 3 (subtask 3)	Traditional	Input: E0003:S003, 這碗拉麵超級無敵霹靂難吃 Output: E0003:S003 (拉麵, 食物#品質, 超級無敵霹靂難吃, 2.00#7.88)
	Simplified	Input: E0003:S003, 这碗拉面超级无敌霹雳难吃 Output: E0003:S003 (拉面, 食物#品质, 超级无敌霹雳难吃, 2.00#7.88)

Table 1: Examples of the dimABSA task

sentiment elements with the corresponding valence-arousal rating values.

The rest of this article is organized as follows. Section 2 provides a description of the Chinese dimensional ABSA shared task. Section 3 introduces the evaluation data construction. Section 4 describes the performance metrics. Section 5 compares evaluation results from the various participating teams. Finally, we conclude this paper with findings and offer future research directions in Section 6.

2 Task Organization

This task aims to evaluate the capability of an automatic system for Chinese dimensional ABSA. The four sentiment elements are defined as follows:

- **Aspect Term** (shorted as **A**):

This denotes an entity indicating the opinion target. If the aspect is omitted without being mentioned clearly, we use "NULL" to represent the term.

- **Aspect Category (C)**

This represents a predefined category for the explicit aspect of the restaurant domain. We use the same categories defined in the SemEval-2016 Restaurant dataset (Pontiki et al., 2016). There are a total of twelve categories; each can be split into an entity and attribute using the symbol "#" as follows: 1) "餐廳#概括" / "餐厅#概括" (restaurant#general); 2) "餐廳#價格" / "餐厅#价格" (restaurant#prices); 3) "餐廳#雜項" / "餐厅#杂项" (restaurant#miscellaneous); 4) "食物#價格" / 食物#价格 (food#prices); 5) "食物#品質" / "食物#品质" (food#quality); 6) "食物#份量與款式" / "食物#份量与款式" (food#style&options); 7) "飲料#價格" / "饮料#价格" (drinks#prices); 8) "飲料#品質" / "饮料#品质" (drinks#quality); 9) "飲料#份量與款式" / "饮料#份量与款式" (drinks#style&options); 10) "氛圍#概括" / "氛围#概括" (ambience#general); 11) "服務#概括" / "服务#概括" (services#general); and 12) "地點#概括" / "地点#概括" (location#general).

- **Opinion Term (O)**

This describes the sentiment words or phrases towards the aspects.

- **Sentiment Intensity (I)**

This reflects sentiments using continuous real-valued scores in the valence-arousal dimensions. The valence represents the degree of pleasant and unpleasant (i.e., positive and negative) feelings, while the arousal represents the degree of excitement and calm. Both the valence and arousal dimensions use a nine-degree scale. Value 1 on the valence and arousal dimensions respectively denotes extremely high-negative and low-arousal sentiment, while 9 denotes extremely high-positive and high-arousal sentiment, and 5 denotes a neutral and medium-arousal sentiment. Valence-arousal values are separated by a hashtag (symbol "#") for a mark.

This dimABSA task can be further divided into three subtasks described as follows.

- **Subtask 1: Intensity Prediction**

The first subtask focuses on predicting sentiment intensities in the valence-arousal dimensions. Given a sentence and a specific aspect, the system should predict the valence-arousal ratings. The input format consists of ID, sentence, and aspect. The output format consists of the ID and valence-arousal predicted values that are separated with a 'space'. The intensity prediction is two real-valued scores rounded to two decimal places and separated by a hashtag, each respectively denoting the valence and arousal rating. Example sentences are presented in Table 1. In Example 1, a given sentence "檸檬醬也不會太油，塔皮對我而言稍軟" (The lemon curd is not too oily and the tart crust is a little soft for me.) and two aspects "檸檬醬" (lemon curd) and "塔皮" (tart crust) as an input, participating systems are expected to respectively predict valence-arousal ratings such as 5.67#5.50 for "檸檬醬" (lemon curd) and 4.83#5.00 for "塔皮" (tart crust).

- **Subtask 2: Triplet Extraction**

The second subtask aims to extract sentiment triplets composed of three elements. Given a sentence only, the system should extract all sentiment triplets (aspect, opinion, intensity). The output format consists of the ID and sentiment triplet that are separated with a 'space'. In Example 2, the input sentence is "不僅餐點美味上菜速度也是飛快耶！！" (The meals were not only delicious but were also served very quickly!!) and the output contains two tuples: the first triple contains "餐點" (meals) as an aspect term, "美味" (delicious) as an opinion term, with valence-arousal ratings as 6.63#4.63; the second triple consists of "上菜速度" (were served) as an aspect term and "飛快" (very quickly) as an opinion term, with valence-arousal ratings as 7.25#6.00.

- **Subtask 3: Quadruple Extraction**

The third subtask aims to extract sentiment quadruples composed of four elements. Given a sentence only, the system should extract all sentiment quadruples (aspect, category, opinion, intensity). The output format consists of the ID and sentiment quadruple that are separated with a 'space'. In Example 3, if the input sentence is "這碗拉麵超級無敵霹靂難吃" (This bowl of ramen is terribly unpalatable.), the expected quadruple includes "拉麵" (ramen) denoted as the aspect which belongs to an aspect category "食物#品質" (food#quality), along with an opinion term "超級無敵霹靂難吃" (terribly unpalatable) and a sentiment intensity value in terms of valence-arousal ratings of 2.00#7.88

3 Data Preparation

We first crawled restaurant reviews from Google Reviews and an online bulletin board system PTT. Then, we removed all HTML tags and multimedia material and split the remaining texts into several sentences. Finally, we randomly selected partial sentences to retain content diversity for manual annotation.

The annotation process was conducted in two phases. We first annotated the aspect/category/opinion elements and then V#A element. In the first phase, three graduate students majoring in computer science were trained to annotate the sentences for aspect/category/opinion. One task organizer led a discussion to clarify annotation differences and seek consensus among the annotators. A majority vote mechanism was finally used to resolve any disagreements among the annotators. In the second phase, each sentence along with the annotated aspect/category/opinion was presented to five annotators majoring in Chinese language for V#A rating. Similarly, one task organizer also led a group discussion during annotation. Once the annotation process was finished, a cleanup procedure was performed to remove outlier values which did not fall within 1.5 standard deviations (SD) of the mean. These outliers were then excluded from calculating the average V#A for each instance.

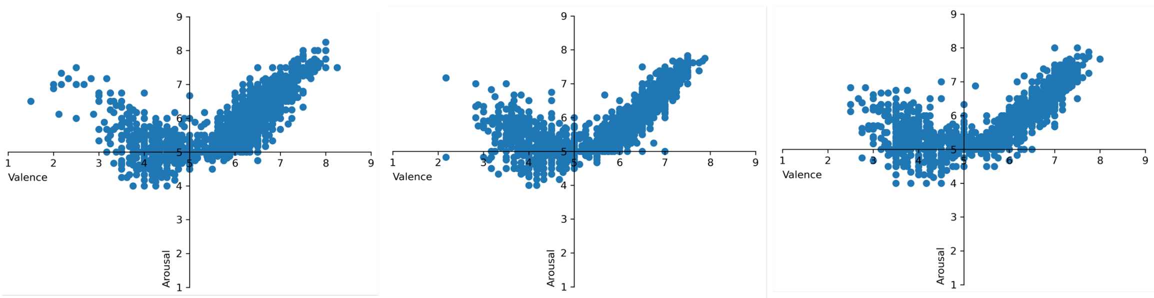
We provided two versions of all datasets with identical content, but one in traditional Chinese characters and the other in simplified Chinese characters. The participating teams could choose their preferred version for the task evaluation. The submitted results were evaluated with the corresponding version of the gold standard and ranked together as the official results.

This shared task is presented as an open test, and participating systems can use other publicly available data, but such data must be specified in the final system description paper. For example, we

Restaurant (REST) Domain									
Subtask	**Dataset**	**#Sent**	**#Char**	**#Tuple**	**Aspect**			**Opinion**	
					#NULL	#Unique	#Repeat	#Unique	#Repeat
ST1	Train	6,050	85,769	8,523	169	6,430	1924	-	-
	Dev.	100	1,109	115	0	115	0	-	-
	Test	2,000	34,002	2,658	0	2,658	0	-	-
ST2 & ST3	Train	6,050	85,769	8,523	169	6,430	1,924	7,986	537
	Dev.	100	1,280	150	0	78	72	143	7
	Test	2,000	39,014	3,566	52	1,693	1,821	3263	303

Table 2: Detailed data statistics

Scatter Plots of Valence-Arousal Distributions



(a) Training Set for Subtasks 1, 2 and 3

(b) Test Set for Subtask 1

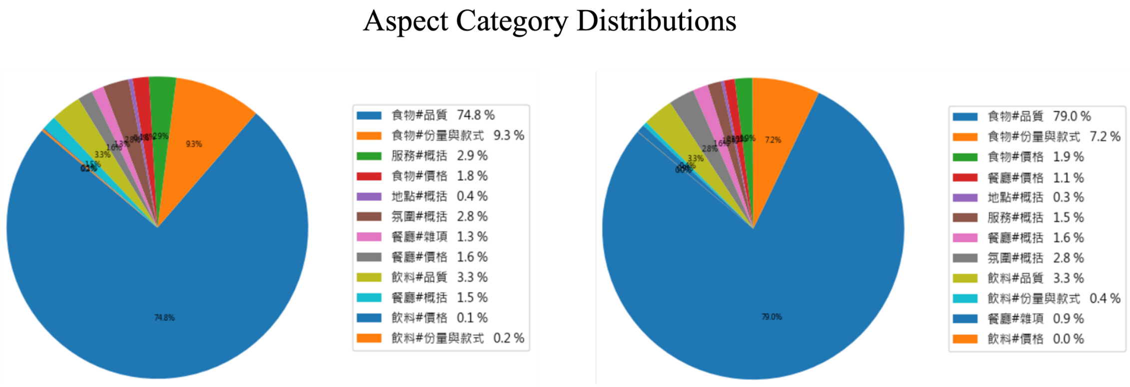
(C) Test Set for Subtasks 2&3

Figure 1: Scatter plots of valence-arousal distributions

also provide the Chinese EmoBank (Lee et al., 2022) as a potentially useful sentiment resource annotated with real-valued scores for both valence and arousal dimensions. This data set features various levels of text granularity including two lexicons called Chinese valence-arousal words (CVAW, 5,512 single words) and Chinese valence-arousal phrases (CVAP, 2,998 multi-word phrases), along with two corpora called Chinese valence-arousal sentences (CVAS, 2,582 single sentences) and Chinese valence-arousal texts (CVAT, 2,969 multi-sentence texts).

Table 1 presents detailed statistics for the mutually exclusive training, development and test sets, where #Sent, #Char, and #Tuple respectively denote the number of sentences, characters and tuples in the dataset. The training set provided for all three subtasks included 6,050 sentences (85,769 characters), annotated with 8,523 tuples. The development set only includes 100 sentences for output format validation. Two mutually exclusive test sets were prepared for system performance evaluation, each including 2,000 sentences. One was provided for Subtask 1 and the other was used for Subtasks 2 and 3.

We further analyzed the aspect types in the test set, including #unique and #repeat which respectively denote the number of aspects which occurred only one time or more than one time. For Subtask 1, a total of 2,658 aspects belong to the unique type, without the null and repeat cases. For Subtasks 2 and 3, 1821 aspects (51.1% out of total 3,566) occurred more than one time across all testing sentences. In addition, a very small portion (near 1.5%) of aspects belonged to the null cases. Similarly, we also analyzed the opinion terms, the repeat cases occupied about 8.5% (=303/3566). These findings revealed: 1) the aspect has a centered distribution, reflecting that users' opinion targets may be similar, and 2) the opinion has a diverse distribution, indicating that different affective words or phrases are used to express a user's feelings.



(a) Training Set for Subtasks 1, 2 and 3

(b) Test Set for Subtask 3

Figure 2: Aspect category distributions

Figure 1 shows the scatter plots of valence-arousal distributions. They presented similar curves for the training and test sets, indicating that both high-positive and high-negative opinion terms usually have high arousal values. Identical results were obtained from the Chinese EmoBank (Lee et al., 2022).

Figure 2 presents the aspect category distributions. The distributions are imbalanced for both the training and test sets for Subtask 3. This finding is the same as that for the SemEval-2016 Restaurant dataset (Pontiki et al., 2016). The most frequently occurring category was "食物#品質" (food#quality), followed by "食物#份量與款式" (food#style&options) and "飲料#品質" (drinks#quality). In the training set, these 3 categories accounted for 87.4% of the total, with the remaining 9 categories accounting for 12.6%. In the test set for Subtask 3, these 3 categories accounted for 89.5% of the total, with the other 9 categories accounting for the remaining 10.5%.

4 Performance Metrics

For Subtask 1, the sentiment intensity prediction performance is evaluated by examining the difference between machine-predicted ratings and human-annotated ratings using two metrics: Mean Absolute Error (MAE) and Pearson Correlation Coefficient (PCC), defined as the following equations.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|a_i - p_i| \tag{1}$$

$$\mathrm{PCC} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{a_i-\mu_A}{\sigma_A}\right)\left(\frac{p_i-\mu_P}{\sigma_P}\right) \tag{2}$$

where $a_i \in A$ and $p_i \in P$ respectively denote the i-th actual value and predicted value, n is the number of test samples, μ_A and σ_A respectively represent the mean value and the standard deviation of A, while μ_P and σ_P respectively represent the mean value and the standard deviation of P.

Each metric for the valence and arousal dimensions is calculated and ranked independently. The actual and predicted real values should range from 1 to 9, so MAE measures the error rate in a range where the lowest value is 0 and the highest value is 8. A lower MAE indicates more accurate prediction performance. The PCC is a value between −1 and 1 that measures the linear correlation between the actual and predicted values. A lower MAE and a higher PCC indicate more accurate prediction performance.

For Subtasks 2 and 3, we use the F1-score as the evaluation metric, defined as:

$$F1 = \frac{\mathrm{2xPxR}}{\mathrm{P+R}} \tag{3}$$

where Precision (P) is defined as the percentage of triplets/quadruples extracted by the system that are correct. Recall (R) is the percentage of triplets/quadruples present in the test set found by the system. The F1-score is the harmonic mean of precision and recall.

Team	Subtask			Architecture		Data Augmentation
	ST1	ST2	ST3	PLM	LLM	
HITSZ-HLT	V	V	V	Erine-3.0-xbase-zg	deepseek-7B-instruct-v1.5	-
CCIIPLab	V	V	V	MacBERT-base	-	Chinese EmoBank
YNU-HPCC	V			BERT-wwm-ext	-	Merged-Train
DS-Group	V			-	GPT-4o	-
TMAK-Plus		V		-	GPT-4o	-
ZZU-NLP		V	V	BERT	Baichuan2-7B	-
JN-NLP			V	-	T5-base	-

Table 3: Participating system summary. ST for subtask, PLM for pre-trained language models, and LLM for large language models.

Subtask 1: Intensity Prediction					
Team	Evaluation Metrics				Overall Rank
	V-MAE	V-PCC	A-MAE	A-PCC	
HITSZ-HLT	**0.279** (1)	**0.933** (1)	**0.309** (1)	**0.777** (1)	**1**
CCIIPLab	0.294 (2)	0.916 (3)	**0.309** (1)	0.766 (3)	2
YNU-HPCC	0.294 (2)	0.917 (2)	0.318 (3)	0.771 (2)	2
DS-Group	0.460 (4)	0.858 (5)	0.501 (4)	0.490 (4)	4
yangnan	1.032 (5)	0.877 (4)	1.095 (5)	0.097 (5)	5

Table 4: Testing results of Subtask 1. V for valence and A for arousal. The best scores of each metrics are in bold.

Each metric for the valence and arousal dimensions is calculated either independently or in combination. First, the valence and arousal values are rounded to an integer. Next, a triplet/quadruple is regarded as correct if and only if the three/four elements and their combination match those in the gold triplet/quadruple. All metrics range from 0 to 1. A higher Precision, Recall, and F1 score indicate more accurate performance.

5 Evaluation Results

5.1 System Summary

We received a total of 214 submissions from 61 registered participants during the evaluation phase. A total of eleven teams provided submissions to the leaderboard for each subtask and seven submitted their task technical papers. HITSZ-HLT (Xu et al., 2024) and CCIIPLab (Tong and Wei, 2024) participated in all three subtasks, ZZ-NLP (Zhu et al., 2024) team took part in two subtasks, and the remaining four teams only joined in one subtask.

Table 3 summarizes the participating systems, including involved subtasks, system architectures and additional data usage. HITSZ-HLT (Xu et al., 2024) integrated a BERT-based pre-trained language model (PLM) (i.e., ERNIE 3.0 (Sun et al., 2021)) and a code-style large language model (LLM) (i.e., deepseek (Guo et al., 2024)) to address this task, demonstrating promising performance in different scenarios. CCIIPLab (Tong and Wei, 2024) proposed a Contrastive Learning-enhanced

Subtask 2: Triplet Extraction				
Team	Evaluation Metrics			Overall Rank
	V-Tri-F1	A-Tri-F1	VA-Tri-F1	
HITSZ-HLT	**0.589** (1)	**0.545** (1)	**0.433** (1)	**1**
CCIIPLab	0.573 (2)	0.522 (2)	0.403 (2)	2
ZZU-NLP	0.542 (3)	0.507 (3)	0.389 (3)	3
BIT-NLP	0.490 (4)	0.450 (4)	0.342 (4)	4
SUDA-NLP	0.475 (5)	0.448 (5)	0.326 (5)	5
TMAK-Plus	0.269 (6)	0.307 (6)	0.157 (6)	6

Table 5: Testing results of Subtask 2. V for valence, A for arousal, VA for valence-arousal, and Tri for triplet. The best scores of each metric are in bold.

Subtask 3: Quadruple Extraction				
Team	Evaluation Metrics			Overall Rank
	V-Quad-F1	A-Quad-F1	VA-Quad-F1	
HITSZ-HLT	**0.567** (1)	**0.526** (1)	**0.417** (1)	**1**
CCIIPLab	0.555 (2)	0.507 (2)	0.389 (2)	2
ZZU-NLP	0.522 (3)	0.489 (3)	0.376 (3)	3
SUDA-NLP	0.487 (4)	0.444 (4)	0.336 (4)	4
JN-NLP	0.482 (5)	0.439 (5)	0.331 (5)	5
BIT-NLP	0.470 (6)	0.434 (7)	0.329 (6)	6
USTC-IAT	0.438 (7)	0.437 (6)	0.312 (7)	7

Table 6: Testing results of Subtask 3. V for valence, A for arousal, VA for valence-arousal, and Quad for quadruple. The best scores of each metric are in bold.

Span-based (CL-Span) framework based on MacBERT (Cui et al., 2021) to improve the performance of tuple extraction and sentiment intensity prediction. The Chinese EmoBank (Lee et al., 2022) was also incorporated as an auxiliary training resource to boost performance. YNU-HPCC (Wang et al., 2024) used a BERT-based encoder to generate aspect-specific representation and train linear predictors to jointly predict valence-arousal ratings. DS-Group (Meng et al., 2024) proposed an aspect-aware example selection method for in-context learning based on LLM. TMAK-Plus (Kang et al., 2024) presented a Multi-Agent Collaboration (MAC) model to assemble several GPT-based LLM for the dimensional ABSA task. ZZU-NLP (Zhu et al., 2024) proposed a two-stage contextual learning approach based on the Baichuan2-7B (Yang et al., 2023). JN-NLP (Jiang et al., 2024) used a paraphrase generation paradigm based on the T5 (Raffel et al., 2020) pre-trained model to address the dimABSA task.

5.2 Official Ranking

Tables 4, 5, and 6 respectively show the testing results for each subtask. Each metric in each individual subtask is ranked independently. (*) means the rank for each metric. A system's overall ranking is computed based on the cumulative rank. The lower the cumulative rank, the better the system performance.

The overall best results came from the HITSZ-HLT (Xu et al., 2024) team, achieving the best scores in all metrics across three subtasks, followed by the CCIIPLab (Tong and Wei, 2024), ranking second on the leaderboard for each subtask.

6 Conclusions and Future Work

This paper provides an overview of the SIGHAN-2024 dimABSA task for Chinese dimensional aspect-based sentiment analysis, including task descriptions, data preparation, performance metrics and evaluation results. We received a total of 214 submissions from 61 registered participants during the evaluation phase. Among eleven participating teams, seven presented their task technical reports. Regardless of actual performance, all submissions contribute to the development of an effective dimensional ABSA solution, and each task technical paper for this shared task also provides useful insights for further research.

We hope the data sets collected and annotated for this shared task can facilitate and expedite future development of Chinese dimensional ABSA. Therefore, the gold standard test set and evaluation scripts are made publicly available in GitHub repositories at: https://github.com/NYCU-NLP/SIGHAN2024-dimABSA

Future directions will focus on the development of Chinese dimensional ABSA models. We plan to build new language resources to develop techniques for the future enrichment of this research topic, especially for reviews in the other domains.

Acknowledgments

We thank all the participants for taking part in our shared task. We appreciate Tzu-Mi Lin, Zue-Yu Xu, Jian-Yu Chou, and all annotators for their efforts in data annotations and verifications. This work is partially supported by the National Science and Technology Council, Taiwan, under the grants NSTC 111-2628-E-A49-029-MY3 and NSTC 113-2221-E-155-046-MY3. The works described in this paper are supported by the National Natural Science Foundation of China (62076158, 62376143), and CCF-Zhipu AI Large Model Foundation of China (CCF-Zhipu202310).

References

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-Category-Opinion-Sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, pages 340-350. https://doi.org/10.18653/v1/2021.acl-long.29

Shaowei Chen, Yu Wang, Jie Liu and Yuelin Wang. 2021. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. The Association for the Advancement of Artificial Intelligence. pages 12666-12674.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504-3514. https://doi.org/10.1109/TASLP.2021.3124365

Yu-Chih Deng, Cheng-Yu Tsai, Yih-Ru Wang, Sin-Horng Chen, and Lung-Hao Lee. 2022. Predicting Chinese phrase-level sentiment intensity in valence-arousal dimensions with linguistic dependency features. *IEEE Access*, 10:126612-126620. https://doi.org/10.1109/ACCESS.2022.3226243

Yu-Chih Deng, Yih-Ru Wang, Sin-Horng Chen, and Lung-Hao Lee. 2023. Towards transformer fusions for Chinese sentiment intensity prediction in valence-arousal domains. *IEEE Access*, 11:109974-109982. https://doi.org/10.1109/ACCESS.2023.3322436

Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao and Weipeng Yan. 2022. LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*. International Committee on Computational Linguistics. pages 7002-7012.

Daya Guo, Qihao Zhu, Dejian Yang, Zhenda Xie, Kai Dong, Wentao Zhang, Guanting Chen, Xiao Bi, Y. Wu, Y. K. Li, Fuli Luo, Yingfei Xiong, and Wenfeng Liang. 2024. DeepSeek-Coder: When the large language model meets programming – the rise of code intelligence. *arXiv preprint*. arXiv:2401.14196. https://doi.org/10.48550/arXiv.2401.14196

Yunfan Jiang, Tianci Liu, and Hengyang Lu. 2024. JN-NLP at SIGHAN-2024 dimABSA task: Extraction of sentiment intensity quadruples based on paraphrase generation. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.

Xin Kang, Zhifei Zhang, Jiazheng Zhou, Yunong Wu, Xuefeng Shi, and Kazuyuki Matsumoto. 2024. TMAK-Plus at SIGHAN-2024 dimABSA task: Multi-agent collaboration for transparent and rational sentiment analysis. In *Proceedings of the*

10th SIGHAN Workshop on Chinese Language Processing. Association for Computational Linguistics.

Lung-Hao Lee, Jian-Hong Li and Liang-Chih Yu. 2022. Chinese EmoBank: Building valence-arousal resources for dimensional sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(4), Article 65, 18 pages. https://doi.org/10.1145/3489141

Yue Mao, Yi Shen, Jingchao Yang, Xiaoying Zhu and Longjun Cai. 2022. Seq2Path: Generating sentiment tuples as paths of a tree. In *Findings of the Association for Computational Linguistics: ACL 2022*. Association for Computational Linguistics, pages 2215-2225. https://doi.org/10.18653/v1/2022.findings-acl.174

Yue Mao, Yi Shen, Chao Yu and Longjun Cai. 2021. A joint training dual-MRC framework for aspect based sentiment analysis. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. The Association for the Advancement of Artificial Intelligence, pages 13543-13551.

Ling-ang Meng, Tianyu Zhao, and Dawei Song. 2024. DS-Group at SIGHAN-2024 dimABSA task: Constructing in-context learning structure for dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu and Luo Si. 2020. Knowing what, how, and why: a near complete solution for aspect-based sentiment analysis. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. The Association for the Advancement of Artificial Intelligence, pages 8600-8607. https://doi.org/10.1609/aaai.v34i05.6383

Joseph J. Peper and Lu Wang. 2022. Generative aspect-based sentiment analysis with contrastive learning and expressive structure. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. Association for Computational Linguistics, pages 6118-6124. https://doi.org/10.18653/v1/2022.findings-emnlp.451

Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud AI-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch. Evgeny Kotelnikov, Nuria Bel, Salud Maria Jimenez-Zafra and Gulsen Eryigit. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, Association for Computational Linguistics, pages 19-30. https://doi.org/10.18653/v1/S16-1002

Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Suresh Manandhar and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, pages 486-495. https://doi.org/10.18653/v1/S15-2082

Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, pages 27-35. https://doi.org/10.3115/v1/S14-2004

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. *Journal of Machine Learning Research*, 21(140): 1-67.

James A Russel. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6): 1161-1178.

Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiaxiang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, Weixin Liu, Zhihua Wu, Weibao Gong, Jianzhong Liang, Zhizhou Shang, Peng Sun, Wei Liu, Xuan Ouyang, Dianhai Yu, Hao Tian, Hua Wu, and Haifeng Wang. 2021. ERNIE 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint*. arXiv:2107.02137. https://doi.org/10.48550/arXiv.2107.02137

Zeliang Tong, and Wei Wei. 2024. CCIIPLab at SIGHAN-2024 dimABSA task: Contrastive learning-enhanced span-based framework for Chinese dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.

Zehui Wang, You Zhang, Jin Wang, Dan Xu, and Xuejie Zhang. 2024. YNU-HPCC at SIGHAN-2024 dimABSA task: Using PLMs with a joint learning strategy for dimensional intensity prediction. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for end-to-end fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for

Computational Linguistics, pages 2576-2585. https://doi.org/10.18653/v1/2020.findings-emnlp.234

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 2339-2349. https://doi.org/10.18653/v1/2020.emnlp-main.183

Hongling Xu, Delong Zhang, Yice Zhang, and Ruifeng Xu. 2024. HITSZ-HLT at SIGHAN-2024 dimABSA task: Integrating BERT and LLM for Chinese dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.

Aiyuan Yang, Bin Xiao, Bingning Wang, Borong Zhang, Ce Bian, Chao Yin, Chenxu Lv, Da Pan, Dian Wang, Dong Yan, Fan Yang, Fei Deng, Feng Wang, Feng Liu, Guangwei Ai, Guosheng Dong, Haizhou Zhao, Hang Xu, Haoze Sun, Hongda Zhang, Hui Liu, Jiaming Ji, Jian Xie, JunTao Dai, Kun Fang, Lei Sun, Liang Song, Lifeng Liu, Liyun Ru, Luyao Ma, Mang Wang, Mickel Liu, MingAn Lin, Nuolan Nie, Peidong Guo, Ruiyang Sun, Tao Zhang, Tianpeng Li, Tianyu Li, Wei Cheng, Weipeng Chen, Xiangrong Zeng, Xiaochuan Wang, Xiaoxi Chen, Xin Men, Xin Yu, Xuehai Pan, Yanjun Shen, Yiding Wang, Yiyu Li, Youxin Jiang, Yuchen Gao, Yupeng Zhang, Zenan Zhou, and Zhiying Wu. 2023. Baichuan 2: Open large-scale language models. *arXiv preprint*, arXiv:2309.10305. https://doi.org/10.48550/arXiv.2309.10305

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 the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, pages 540-545. https://doi.org/10.18653/v1/N16-1066

Li Yuan, Jin Wang, Liang-Chih Yu, Xuejie Zhang. 2023. Encoding Syntactic Information into Transformers for Aspect-Based Sentiment Triplet Extraction. *IEEE Transactions on Affective Computing*, 15(2): 722-735. https://doi.org/10.1109/TAFFC.2023.3291730

Chen Zhang, Qiuchi Li, Dawei Song, and Benyou Wang. 2020. A multi-task learning framework for opinion triplet extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, pages 819-829. https://doi.org/10.18653/v1/2020.findings-emnlp.72

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing and Wai Lam. 2021. Towards generative aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Short Papers)*. Association for Computational Linguistics, pages 504-510. https://doi.org/10.18653/v1/2021.acl-short.64

Junxian Zhou, Haiqin Yang, Yuxuan He, Hao Mou. and Junbo Yang. 2023. A unified one-step solution for aspect sentiment quad predication. In *Findings of the Association for Computational Linguistics: ACL 2023*. Association for Computational Linguistics, pages 12249-12265. https://doi.org/10.18653/v1/2023.findings-acl.777

Senbin Zhu, Hanjie Zhao, Xingren Wang, Shanhong Liu, Yuxiang Jia, and Hongying Zan. 2024. ZZU-NLP at SIGHAN-2024 dimABSA task: Aspect-based sentiment analysis with coarse-to-fine contextual learning. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.