Conversational Aspect-Based Sentiment Quadruple Analysis with Consecutive Multi-view Interaction

Yongquan Lai[2,3], Shixuan Fan[1,2], Zeliang Tong[1,2], Weiran Pan[1,2], and Wei Wei[1,2(✉)]

[1] Cognitive Computing and Intelligent Information Processing (CCIIP) Laboratory, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China
`{fanshixuan,tongzeliang,panwr,weiw}@hust.edu.cn`
[2] Joint Laboratory of HUST and Pingan Property and Casualty Research (HPL), Wuhan, China
[3] Ping An Property and Casualty Insurance Company of China, Ltd., Wuxi, China
`laiyongquan932@pingan.com.cn`

Abstract. With the development of information technology, the increasing amount of content on the web has made aspect-based sentiment analysis an essential tool for extracting information about emotional states. However, most of the existing work focuses on a single text, while little attention is paid to the task of sentiment analysis in complex texts such as dialogues, in which the quadruple of target-aspect-opinion-sentiment may appear in different speakers during one conversation thread. In this paper, we proposed a novel framework that was built by a Chinese pre-trained language model and a grid tagging classifier. In addition, we use multi-view interaction with three consecutive multi-head attention modules to improve the performance and robustness of our model. Besides, based on the excellent performance of the Chinese pre-training model, the English version is transferred from the final Chinese weights to achieve cross-lingual transfer. To improve the generalization ability of the model, cross-validation is used to select the best one. Our model ranks first on track 4 of the NLPCC-2023 shared task on conversational aspect-based sentiment quadruple analysis. Our code is publicly available at https://github.com/Joint-Laboratory-of-HUST-and-PAIC/nlpcc2023-shared-task-diaASQ.

Keywords: Pre-trained language model · Attention · Cross-validation

1 Introduction

Sentiment analysis is one of the most active research areas in natural language processing [11]. In recent years, due to the surge of user feedback information

---

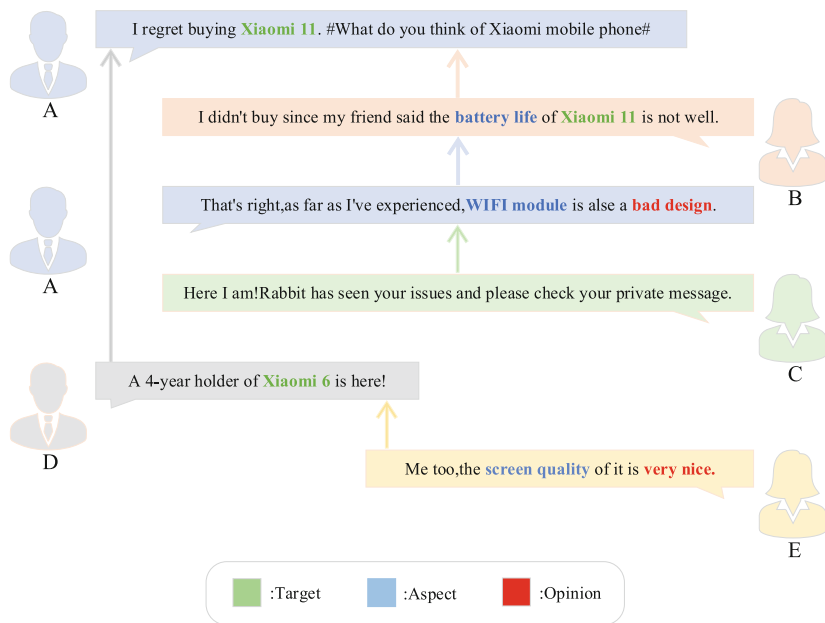Y. Lai and S. Fan—Contribute equally to this paper.

Fig. 1. An example data of the conversational aspect-based sentiment quadruple analysis (DiaASQ). In this dialogue, three quadruples appear, which are ('*Xiaomi 11*', '*WIFI module*', '*bad design*', '*negative*'), ('*Xiaomi 11*', '*battery life*', '*not well*', '*negative*') and ('*Xiaomi 6*', '*screen quality*', '*very nice*', '*positive*').

about goods, social events, news, and other content on social media, the research of this task has attracted wide attention [16]. In linguistics, the sentiment elements of text are mainly composed of several parts: aspect term, aspect entity, opinion term, and sentiment polarity [19]. Based on the above linguistic background, relevant researchers have formally proposed a formal definition of aspect-based sentiment analysis (ABSA) [4], which the sentiment polarity was directed by an entity or aspect, not the whole sentence. Because different tasks have different combinations of elements extracted, the ABSA task was divided into a number of subtasks with different goals [21].

Compared to the single text piece, dialogue text usually contains more information, including the background, context, characteristics and relationship of the speaker [10], which are all unique challenges in the task of dialogue sentiment analysis and have a huge influence on emotional tendencies. Therefore, there is an urgent need to design a conversational aspect-based sentiment quadruple analysis framework, which aims to detect the fine-grained sentiment quadruple of target-aspect-opinion-sentiment. In this task, for a given data, our goal is to extract all the sentiment quadruples appeared in the dialogue. For example in Fig. 2, there are three sentiment quadruples that appeared in the dialogue. Although some scholars have made great efforts in the past, there is still a large degree of distortion in conversational aspect-based sentiment quadruple extrac-

tion [13]. First, it is always difficult to model the characteristics of multi-person conversations in a multi-person conversation. Different people may express sentiment on aspect terms at different times, and there may be a time delay between sentiment expressions. So this non-synchronicity makes sentiment analysis more challenging, which requiring consideration of emotional interaction and evolution at different points in time. Second, in a multi-person conversation, the emotions of different people may influence and compete with each other. Therefore, it's significant for us to analyze the interaction between multiple people and the impact of emotions accurately. Third, exactly locating the boundary of emotional elements and linking the same set of emotional elements are also one of the research difficulties.

To address those issues, Li et al. proposed DiaASQ [9], an end-to-end neural model for sentiment quadruple analysis. DiaASQ conducts feature interaction at three different views (i.e., speaker, reply, and thread) independently, and uses the max-pooling to aggerate information from different views. We argue this independent multi-view interaction manner may not utilize the dialogue information sufficiently. For example, when conducting thread-level interaction, DiaASQ ignores the reply relation between utterances and the model is not aware of which utterances come from the same speaker. Such shortcomings limited the model to comprehensively capturing the emotional interaction and evolution during the dialogue.

To this end, we introduce a novel multi-view interaction module consisting of three consecutive multi-head attention layers. Specifically, we first conduct the feature interaction between tokens from the same speaker to model the emotional state of each speaker. Then we conduct feature interaction between utterances and their corresponding replies to model the local emotional interaction. Finally, we allow tokens in the same thread to interact with each other to generate dialogue-specific features. This hierarchical feature interaction architecture allows us to aggregate emotional information from the local single speaker to the global multi-round dialogue and we experimentally proved this method brings a considerable performance improvement.

We also utilize other modules to further improve the model performance. To better adapt the Chinese data, we use macbert [3] as the encoder, which is pre-trained on a large corpus of Chinese text. And the English version model is transferred from the final Chinese weights to achieve cross-lingual transfer. In addition, we use k-fold validation to select the best models and ensemble them by weight averaging. The main contributions of this work are summarized as follows:

1. We deployed hierarchical feature interaction from the three levels of speaker, reply, and thread successively, and carries out multi-granularity feature interaction from local to global.
2. We use macbert as the encoder in order to better adapt to Chinese data. At the same time, the English model is fine-tuned on the basis of the Chinese model, and the effect is improved through cross-language interaction.

3. We experimental show that our method achieves state-of-the-art results on the conversational aspect-based sentiment quadruple analysis. The ablation study also proves the effectiveness of each component.

2   Related Works

2.1   Aspect Sentiment Triplet Extraction

As a compound ABSA task, the aspect sentiment triplet extraction (ASTE) task attempts to extract sentiment triplets from a given sentence that tell us what the opinion goal is, what its emotional tendency is, and why that emotion is expressed via opinion terms. Researchers have done several valuable attempts in ASTE task. Peng et al. [12] first proposed a two-stage pipeline model to extract the triplet of sentiment elements, which extract sentiment elements and construct aspect-opinion pairs separately. However, the pipeline method ignore the interaction between sentiment elements and commonly suffer from error propagation. So Wu et al. [17] extend the grid tagging scheme (GTS) applied to other ABSA task to predict sentiment triplets, while this methods rely on the interaction between word pairs. Xu et al. [18] proposed a span-level interaction model that explicitly considers the interaction between the span of the entire Aspect Term and Opinion Term. Their approach significantly improves the performance, especially on sentiment triplets which contain multi word targets or opinions. In order to further improve the model effect, Chen et al. [2] design a span-level bidirectional network, which includes a span separation loss to ensure that spans containing shared tokens have distinct representations.

Unlike the task that needs to be solved in this article, the ASTE task blurs the boundary between aspect term and aspect entity, and this article needs to accurately distinguish the two based on their differences.

2.2   Emotion Recognition in Conversation

As an extension of the basic task of sentiment analysis, conversational sentiment analysis has attracted wide attention in the field of natural language processing, and many researchers have focused on related research work [13]. Because dialogue is a dynamic process in which the emotional expression between participants influences and evolves. Therefore, it is necessary to focus on the emotional interaction of the participants in the dialogue and the evolution of emotions, rather than simply viewing the dialogue separately. Hazarika et al. [7] propose a conversational memory network that incorporates audio, visual, and textual features to capture dependencies between speakerstried and model the historical conversation information of people in the conversation. Ghosal et al. [6] proposed a graph-based convolutional neural network for emotion recognition in conversation, which used the relationship graph in the conversation to simulate the propagation and influence of emotion. Hu et al. [8] designed multiple rounds of reasoning modules to extract and integrate emotional cues, which fully understand the conversational context from a cognitive perspective.
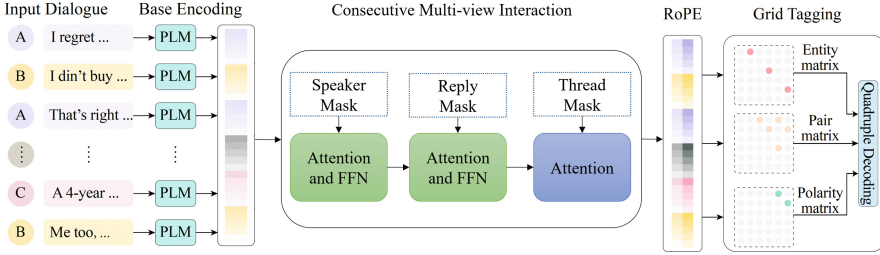
Fig. 2. Architecture of the model. The input dialogue for both languages are encoded by the same encoder, namely Macbert-large [3], and the output of the encoder is fed into three consecutive attention modules, as shown in the figure. The first two attention modules contain multi-head attention and feed-forward network (FFN), and they capture speaker and reply relations using speaker mask and reply mask respectively.

The task of emotion recognition in conversation needs to fully consider the context information, which can reveal the transfer process of emotion in dialogue by modeling the context, that is, the transmission and influence of emotion. The method mentioned above can be widely applied in this task to better distinguish different aspects of emotion and improve the accuracy of conversational aspect-based sentiment quadruple extraction.

3   Methodology

Our model is an improvement over the DiaASQ model [9], and the architecture is shown in Fig. 1.

3.1   Task Definition

Given a multi-user dialogue context $\mathcal{D} = \{u_1, \cdots, u_n\}$ with the corresponding replying record $l = \{l_1, \cdots, l_n\}$ of utterances, where l_i denotes that i-th utterance response to l_i utterance and each utterance u_i is composed by m length words w_j, denoted as $u_i = \{w_1, \cdots, w_m\}$. Based on $\mathcal{D}$ and l, the ABSA task aims to extract all target-aspect-opinion-sentiment quadruples, denoted as $Q = \{t, a, o, p\}_{k=1}^{K}$, where the target t, aspect a, opinion o are a sub-string of the dialogue context, separately, and the sentiment p is a category label $\in \{pos, neg, other\}$.

Following previous work [9], we split the ABSA task into three subtasks, namely *entity boundary, entity pair* and *sentiment polarity*. For *entity boundary* subtask, we use *tgt*, *asp*, and *opi* label to mark the head and tail of target, aspect, and opinion in the dialogue context. *Entity pair* subtask aims to use *h2h* (head-to-head) and *t2t* (tail-to-tail) labels to link different types of terms together as a combination (t, a, o). *Sentiment polarity* subtask is a sentiment classification task and we assign the category label (*i.e. pos, neg, other*) between the heads and tails of target and opinion terms.
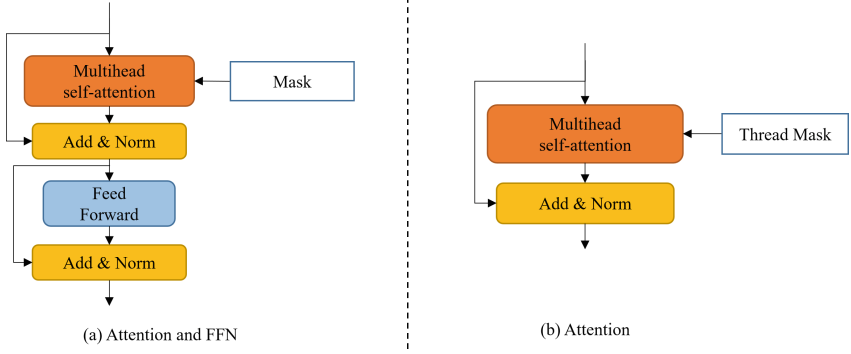
(a) Attention and FFN

(b) Attention

Fig. 3. Attention modules used in our method. The attention module is composed of multi-head attention and feed-forward network (FFN). The multi-head attention is the same as the one in the transformer [15], and the FFN (for the first two attention modules in 1) is a two-layer MLP with GELU activation. (Color figure online)

3.2 Base Encoding

We use macbert [3] as the pretrained language model (PLM) for both English and Chinese encoders. The macbert modifies the masked language model (MLM) task as a language correction manner to mitigate the discrepancy of the pre-training and fine-tuning stage. The output of the last attention module is fed into the RoPE and grid tagging modules as in the DiaASQ implementation [9].

$$u_i^{'} = < [\text{CLS}], w_1, \cdots, w_m, [\text{SEP}] >, \tag{1}$$

$$\boldsymbol{H_i} = \boldsymbol{h}_{cls}, \boldsymbol{h}_1, \cdots, \boldsymbol{h}_m, \boldsymbol{h}_{sep} = \text{PLM}(u_i^{'}), \tag{2}$$

where $\boldsymbol{h}_m$ is the contextual representation of word w_m.

3.3 Consecutive Multi-view Interaction

In order to capture the deep interaction between different views, we design a consecutive multi-view interaction module, which captures the correlation in different views respectively through three consecutive attention layers. The first two attention modules consist of multi-head attention and feed-forward network(FFN) [15], shown in Fig. 3(a), capturing speaker and reply relations by using speaker mask and reply mask respectively.

$$\boldsymbol{H}' = \text{Masked-Att}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}, \boldsymbol{M}^o)$$
$$= \text{Softmax}(\frac{(\boldsymbol{Q}^T \cdot \boldsymbol{K}) \odot \boldsymbol{M}^o}{\sqrt{d}}) \cdot \boldsymbol{V}, \tag{3}$$

$$\boldsymbol{H}^o = \text{FFN}(\boldsymbol{H}')$$
$$= \max(0, \boldsymbol{H}'W_1 + b_1)W_2 + b_2, \tag{4}$$

where $\boldsymbol{Q}=\boldsymbol{K}=\boldsymbol{V}=\boldsymbol{H} \in \mathbb{R}^{N \times d}$ is the whole dialogue sequence representation obtained by pre-trained language model, and $\odot$ is element-wise production. The o represent the token interaction under different views respectively, *i.e.* speaker, reply, and thread.

The last attention module (in blue) contains a multi-head attention only, shown in Fig. 3(b), as we found that adding FFN degrades the performance (see ablation study), and thread mask is used to capture the thread relation. We denote the dialog representation after continuous multi-view interaction as $\boldsymbol{H}^f$.

3.4   Quadruple Decoding

First, we utilize multiple MLP layers with unshared parameters to map the dialogue context representations into multiple tag spaces respectively.

$$\boldsymbol{v}_i^r = \mathrm{MLP}^r(\boldsymbol{h}_i^f), \tag{5}$$

where $r \in \{tgt, \cdots, h2h, \cdots, pos, \cdots, \epsilon_{ent}, \cdots\}$ indicates a specific label and ϵ_{ent} denotes the non-relation label in the entity boundary matrix.

In order to help the model understand the dialogue context order, following previous work [9], we fuse the rotation position embedding (RoPE) [14] with the dialogue representation as the input of the quadruple decoding. RoPE can model the relative positional distance between tokens, which can be formalized as follows:

$$\boldsymbol{h}_i^r = \boldsymbol{\mathcal{R}}(\theta, i)\boldsymbol{v}_i^r, \tag{6}$$

where $\boldsymbol{\mathcal{R}}(\theta, i)$ is a positioning matrix parameterized by θ and the absolute index i of $\boldsymbol{v}_i^r$.

For entity boundary subtask, we compute the dot product similarity between tokens as a label score s_{ij}^r, and use softmax to compute the probabilities p_{ij}^r of multiple labels. Other subtasks also get label probabilities through the same method.

$$s_{ij}^r = (\boldsymbol{h}_i^r)^T \boldsymbol{h}_j^r,$$
$$p_{ij}^{ent}, p_{ij}^{tgt}, p_{ij}^{asp}, p_{ij}^{opi} = \mathrm{Softmax}([s_{ij}^{\epsilon_{ent}}; s_{ij}^{tgt}; s_{ij}^{asp}; s_{ij}^{opi}]), \tag{7}$$

During model training, we use the commonly used cross-entropy loss function for each subtask.

$$\mathcal{L}_k = -\frac{1}{G \cdot N^2} \sum_{g=1}^{G} \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{\alpha}^k \, y_{ij}^k \log(p_{ij}^k), \tag{8}$$

where $k \in \{ent, pair, pol\}$ indicates one of subtasks, and $\boldsymbol{\alpha}^k$ is the label weight to alleviate the problem of label imbalance in the dataset. The final loss $\mathcal{L}$ is the weighted sum of three subtask losses:

$$\mathcal{L} = \mathcal{L}_{ent} + \boldsymbol{\beta}\mathcal{L}_{pair} + \boldsymbol{\eta}\mathcal{L}_{pol}, \tag{9}$$

where $\mathcal{L}_{ent}, \mathcal{L}_{pair}$, and $\mathcal{L}_{pol}$ represent the losses of the three subtasks, namely entity boundary, entity pair and sentiment polarity, separately. $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are specify hyperparameters.

Table 1. Data statistics.

		Dialogue				Items		Pairs			Quadruples		
		Dia.	Utt.	Spk.	Tgt.	Asp.	Opi.	$Pair_{t-a}$	$Pair_{t-o}$	$Pair_{a-o}$	Quad.	Intra.	Cross
ZH	Total	1,000	7,452	4,991	8,308	6,572	7,051	6,041	7,587	5,358	5,742	4,467	1,275
	Train	800	5,947	3,986	6,652	5,220	5,622	4,823	6,062	4,297	4,607	3,594	1,013
	Valid	100	748	502	823	662	724	621	758	538	577	440	137
	Test	100	757	503	833	690	705	597	767	523	558	433	125
EN	Total	1,000	7,452	4,991	8,264	6,434	6,933	5,894	7,432	4,994	5,514	4,287	1,227
	Train	800	5,947	3,986	6,613	5,109	5,523	4,699	5,931	3,989	4,414	3,442	972
	Valid	100	748	502	822	644	719	603	750	509	555	423	132
	Test	100	757	503	829	681	691	592	751	496	545	422	123

3.5   Training Strategy

Cross Validation and Model Fusion. We randomly split the training data into 5 folds, and train the model on each fold. For each fold, we select the best model on the validation set and use the weight for model fusion. The model fusion is done by averaging the weights of the selected models. Note that we only select the top 3 models among the 5 folds, as we found that using more models degrades the performance.

Language Transfer from Chinese to English. We found that transfer learning from Chinese to English is effective, and we use the following method to transfer the model from Chinese to English. First, we train the model on Chinese data using cross-validation and model fusion(as described above). Then we use the fused Chinese model to initialize the parameters of the English model. Finally, the English model is trained as usual. We found that this method is more effective than training the model from scratch on English data.

Sentiment Correction by Rules. We extract the apsect-opinion pairs from the training set and build the rules based on the pairs. We keep the top 512 pairs both positive and negative sentiments. We remove the pairs that appear in both positive and negative sentiments, resulting in 426 pairs for both negative and postive and both. The numbers of pairs for both languages are the same probably because the english dataset is directly translated from the chinese dataset. These pairs are used to correct the sentiment prediction in a simple manner: if the pair appears in the prediction, we change the corresponding sentiment to the sentiment of the pair.

4   Experiment

4.1   Experimental Settings

Dataset. DiaASQ [9] is a mobile phone field conversational aspect-based senti-
ment quadruple analysis dataset, collected from Weibo. Each conversation orig-
inates from a root post, and multiple speakers participate in replying to pre-
decessor posts. Multiple threads and multiple turns of conversation form a tree
structure. The statistical information of the dataset is shown in Table 1.

Table 2. Main results on the offline test set. The best results are in bold.

Method	Chinese		English		Avg. F1
	Micro F1	Iden.F1	Micro F1	Iden.F1	
CRF-Extract-Classify	8.81	9.25	11.59	12.80	10.61
SpERT	13.00	14.19	13.07	13.38	13.41
ParaPhrase	23.27	27.98	24.54	26.76	25.64
Span-ASTE	27.42	30.85	26.99	28.34	28.40
DiaASQ	34.94	37.51	33.31	36.80	35.64
Ours	**43.00**	**46.20**	**38.32**	**44.06**	**42.90**

Evaluation Metrics. Following previous works [9], we evaluate all methods on
quadruple extraction, using *micro F1* and *identification F1* as metrics respec-
tively. *Micro F1* measures the whole quadruple, including the sentiment polarity
and *identification F1* does not distinguish the polarity.

Alternative Baselines. Following previous work [9], we select **CRF-Extract
Classify** [1], **SpERT** [5], **Span-ASTE** [18], **ParaPhrase** [20] and **DiaASQ** [9]
as baselines, where DiaASQ is the official baseline of the track 4 of the NLPCC-
2022 shared task.

Implementation Details. We take the Chinese-macbert-large [3] as the pre-
train language model for the Chinese and English datasets. Throughout the
experiments, we use Adam optimizer, where the initial learning rate is 1e-6. In
order to prevent overfitting, the dropout rate is fixed at 0.2. Specify hyperpa-
rameters β and η are set to 3.

4.2   Main Comparisons

All evaluation results under automatic metrics are reported in 2. We can observe
that, among all models, our method achieves the best results in evaluation met-
rics. Compared with the official baseline, our method improves by 8.06%, 8.69%,
5.01%, and 7.26% on the micro F1 and identification F1 evaluation metrics for

Chinese and English datasets, respectively. Compared with the official baseline model (DiaASQ), the performance improvement of our solution mainly comes from a stronger pre-training model, more sufficient feature interaction, and our training strategy. Our approach also achieves the best result in the DiaASQ competition.

Table 3. Ablation results

	Chinese		English		Avg. F1
	Micro-F1	Iden-F1	Micro-F1	Iden-F1	
DiaASQ	34.94	37.51	33.31	36.80	35.64
w/o PLM	14.21	17.55	15.68	19.57	16.75
Ours	43.00	46.20	38.32	44.06	42.90
w/o PLM	17.52	21.33	16.97	20.28	19.03
w DiaASQ PLM	40.46	42.80	35.76	41.65	40.17
w max pooling	41.79	43.55	36.05	41.20	40.65
w all ffn	42.19	44.68	36.57	42.27	41.43
w/o k-fold	41.21	43.12	36.35	42.72	40.85
w/o trans	41.22	43.87	36.83	41.43	40.84
w/o rule	42.82	45.93	37.89	43.77	42.60

4.3   Ablation Study

In order to verify the effectiveness of each optimization we made to the official baseline (DiaASQ), we conducted detailed comparison experiments. The experimental results are shown in Table 3.

For **pre-trained language model**, we design two variants: removing the pre-trained model entirely, using only randomly initialized word2vec (w/o PLM), and using a pre-trained language model consistent with the baseline model (w DiaASQ PLM). We can observe that different pre-trained language models have a greater impact on performance, and the version without PLM drops significantly, proving the important role of pre-trained language models in modeling semantic relevance.

For **methodology**, we replaced the multi-view deep interaction module with multi-view interaction max-pooling in the official baseline model (w max pooling), and the average F1 on the Chinese and English datasets dropped by 1.93% and 2.56%, respectively. This proves that multi-view deep interaction has a strong ability to aggregate information from different views, which is beyond the reach of the max pooling in the baseline method. We also add ffn layer for the thread view in consecutive multi-view interaction module (w all ffn), and the average F1 dropped by 0.47%.

For **training strategy**, we remove k-fold model fusion (w/o k-fold), language transfer (w/o trans), and rule-based sentiment correction (w/o rule), separately.

The performers of three variants of our method decrease by 2.05%, 2.06% and 0.29% in the average F1 metric respectively, which demonstrates the importance of model ensemble and cross-lingual learning to further improve model performance and stability.

5   Conclusion

In this paper, we use a Chinese pre-trained language model and grid tagging schema as backbone to tackle the problem of conversational aspect-based sentiment quadruple analysis. We deploy a multi-view interaction module consisting of three consecutive multi-head attention layers to aggregate emotional information from the local single speaker to the global multi-round dialogue. Besides, the English version model is transferred from the final Chinese weights and k-fold validation is used to improve the model performance. Finally, our proposed framework received second place in the NLPCC 2023 task 4, with an average F1 score of 42.89%. Although exciting improvements over the baseline DiaASQ model appear, our results show that the conversational aspect-based sentiment quadruple analysis is still challenging, which needs further consideration and discussion.

References

1. Cai, H., Xia, R., Yu, J.: 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 (Volume 1: Long Papers), pp. 340–350 (2021)

2. Chen, Y., Keming, C., Sun, X., Zhang, Z.: A span-level bidirectional network for aspect sentiment triplet extraction. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pp. 4300–4309. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, December 2022. https://aclanthology.org/2022.emnlp-main.289

3. Cui, Y., Che, W., Liu, T., Qin, B., Wang, S., Hu, G.: Revisiting pre-trained models for Chinese natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings. pp. 657–668. Association for Computational Linguistics, Online, November 2020. https://www.aclweb.org/anthology/2020.findings-emnlp.58

4. Do, H.H., Prasad, P.W., Maag, A., Alsadoon, A.: Deep learning for aspect-based sentiment analysis: a comparative review. Expert Syst. Appl. **118**, 272–299 (2019)

5. Eberts, M., Ulges, A.: Span-based joint entity and relation extraction with transformer pre-training. In: ECAI 2020, pp. 2006–2013. IOS Press (2020)
6. Ghosal, D., Majumder, N., Poria, S., Chhaya, N., Gelbukh, A.: Dialoguegcn: a graph convolutional neural network for emotion recognition in conversation. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 154–164 (2019)
7. Hazarika, D., Poria, S., Zadeh, A., Cambria, E., Morency, L.P., Zimmermann, R.: Conversational memory network for emotion recognition in dyadic dialogue videos. In: Proceedings of NAACL-HLT, pp. 2122–2132 (2018)
8. Hu, D., Wei, L., Huai, X.: Dialoguecrn: contextual reasoning networks for emotion recognition in conversations. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 7042–7052 (2021)
9. Li, B., et al.: Diaasq: a benchmark of conversational aspect-based sentiment quadruple analysis. arXiv preprint arXiv:2211.05705 (2022)
10. Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., Cambria, E.: Dialoguernn: an attentive RNN for emotion detection in conversations. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 6818–6825 (2019)
11. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**(1–2), 1–135 (2008)
12. Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., Si, L.: Knowing what, how and why: a near complete solution for aspect-based sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 8600–8607 (2020)
13. Poria, S., Majumder, N., Mihalcea, R., Hovy, E.: Emotion recognition in conversation: research challenges, datasets, and recent advances. IEEE Access **7**, 100943–100953 (2019)
14. Su, J., Lu, Y., Pan, S., Wen, B., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding. CoRR abs/2104.09864 (2021), arxiv.org/abs/2104.09864
15. Vaswani, A., et al.: Attention is all you need. Advances in neural information processing systems 30 (2017)
16. Vinodhini, G., Chandrasekaran, R.: Sentiment analysis and opinion mining: a survey. Int. J. **2**(6), 282–292 (2012)
17. Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., Xia, R.: Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In: Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 2576–2585 (2020)
18. Xu, L., Chia, Y.K., Bing, L.: Learning span-level interactions for aspect sentiment triplet extraction. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 4755–4766 (2021)
19. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: a survey. Wiley Interdisciplinary Rev. Data Mining Knowl. Discovery **8**(4), e1253 (2018)
20. Zhang, W., Deng, Y., Li, X., Yuan, Y., Bing, L., Lam, W.: Aspect sentiment quad prediction as paraphrase generation. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 9209–9219 (2021)
21. Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W.: A survey on aspect-based sentiment analysis: tasks, methods, and challenges. IEEE Trans. Knowl. Data Eng. (2022)