MLDSP-MA: Multidimensional Attention for Multi-Round Long Dialogue Sentiment Prediction

Yunfei Yin*, Congrui Zou*, Zheng Yuan*, Xianjian Bao*
*College of Computer Science, Chongqing University, Chongqing 400044, China
yinyunfei@cqu.edu.cn, zoucr@alu.cqu.edu.cn, yuanzh@cqu.edu.cn, xibao@mum.edu

Abstract
The intelligent chatbot takes dialogue sentiment prediction as the core, and it has to tackle long dialogue sentiment prediction problems in many real-world applications. Current state-of-the-art methods usually employ attention-based dialogue sentiment prediction models. However, as the conversation progresses, more topics are involved and the changes in sentiments become more frequent, which leads to a sharp decline in the accuracy and efficiency of the current methods. Therefore, we propose a Multi-round Long Dialogue Sentiment Prediction based on Multidimensional Attention (**MLDSP-MA**), which can focus on different topics. In particular, **MLSDP-MA** leverages a sliding window to capture different topics and traverses all historical dialogues. In each sliding window, the contextual dependency, sentiment persistence, and sentiment infectivity are characterized, and local attention cross fusion is performed. To learn dialogue sentiment globally, global attention is proposed to iteratively learn comprehensive sentiments from historical dialogues, and finally integrate with local attention. We conducted extensive experimental research on publicly available dialogue datasets. The experimental results show that, compared to the current state-of-the-art methods, our model improves by 3.5% in accuracy and 5.7% in Micro-F1 score.

Keywords: dialogue sentiment, chatbots, sliding windows, global attention, local attention

1. INTRODUCTION

Unlike traditional dialogue sentiment prediction, multi-round long dialogue sentiment prediction typically involves more than twenty sentences (Wen et al., 2023; Zhou et al., 2023). This characteristic of multi-round long dialogue sentiment prediction poses new challenges. For example, in dialogues between intelligent chatbots (Huang et al., 2023; Calabrese et al., 2023) and customers, which can last for several hours, dozens of conversation topics may be involved. If traditional dialogue sentiment prediction methods are adopted, the sentimental dynamics of the conversation cannot be accurately captured. This is because traditional dialogue sentiment prediction only considers sentiment prediction for a single topic and does not take into account topic changes. If the conversation topic changes, the corresponding sentiment may also change. For instance, both parties in the conversation may have positive sentiment towards the "travel" topic, but negative sentiment towards the "weather" topic. If the sentiment prediction for the new topic is based on the sentiment of the previous topic, errors may occur. Multi-round long dialogues involve multiple conversations and multiple topics. Frequent topic changes require corresponding sentimental changes to adapt. Therefore, multi-round long dialogue sentiment prediction faces the challenge of multiple topic changes.

Given those challenges, the academic community has proposed sentiment learning-based prediction methods and advocated for the construction of effective neural network models to predict dialogue sentiment. For example, Bothe et al. (2017) used long short-term memory (LSTM) networks to synthesize the influence of historical dialogues on current dialogue sentiment. Building on such research, Cai et al. (2020) introduced sentiment classification and emotion behavior classification methods to further refine the factors affecting dialogue sentiment, thereby improving the accuracy of dialogue sentiment prediction. Recently, attention-based dialogue sentiment prediction has received great attention from scholars, where pair-wise and seq-wise relationships (Wang et al., 2020) were effective methods that proposed a neural simulation model to simulate the next sentiment. To delve into the factors that influence current dialogue sentiment, Zou et al. (2022) discussed the distinction between emotion and sentiment and extracted terms such as joy, surprise, trust, anticipation, anger, disgust, concern, and sadness from dialogues to enhance the accuracy of dialogue sentiment prediction.

Although proven to be effective, the methods based on sliding window for capturing contextual dependency, sentiment persistence, and sentiment infectivity have not been mentioned, which is a potential method to improve the accuracy of sentiment prediction in dialogue. For example, as shown in Figure 1, with the increase of historical dialogue, the topics involved also increase. If the historical topics cannot be effectively sorted out and distinguished, it will affect the efficiency and accuracy of the final sentiment prediction in dialogue. To alleviate this problem, we propose a multi-dimensional attention-based multi-round long dialogue sentiment prediction model, **MLDSP-MA**. In particular, the sliding window technique is first leveraged to capture different topics. In each topic, contextual dependency, sentiment persistence, and sentiment infectivity are proposed to assist in predicting the sentiment of the dialogue. Furtherly, the local attention cross fusion mechanism of **MSDP-MA** merges contextual dependency, sentiment persistence, and sentiment contagion. To learn the sentiment in dialogue from a global perspective, the global attention fusion mechanism of **MSDP-MA** is proposed to exploit all the historical dialogue data to learn the sentiment in dialogue from a global perspective. It is essentially an attention-based decoder, as shown in Figure 2. Experimental results

LREC-COLING 2024, pages 11405–11414

show that our method outperforms the current state-of-the-art methods.

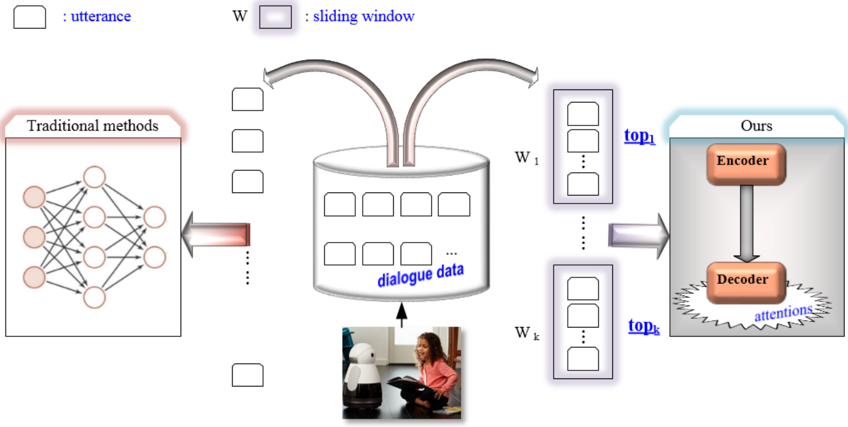

Figure 1: Topic change issue in sentimental prediction of multi-round long conversations. In Figure 1, on the left side, it is the existing methods for predicting dialogue sentiment, which learn the dialogue sentiment from all historical dialogues. Therefore, when encountering multi-round long dialogue scenarios, those methods not only have low accuracy but also low efficiency. On the right side of Figure 1, it is our **MSDP-MA**, which proposes a multi-dimensional attention approach using the sliding window technique. Experimental results show that it outperforms the existing state-of-the-art methods.

In summary, the main contributions of this paper can be summarized as follows.

(1) The sliding window is first introduced to capture the statements/utterances within a topic, focusing on the current topic while traversing all historical dialogues. This provides a method to extend dialogue sentiment prediction to multi-round long dialogue sentiment prediction.

(2) Multi-dimensional attention-based sentiment prediction is first proposed, which takes the pairwise interaction of contextual dependency, sentiment persistence, and sentiment infectivity within the sliding window as local attention, and the comprehensive sentiment iteratively trained from historical dialogue data as global attention. This provides a dialogue sentiment prediction method that combines local attention and global attention.

(3) We demonstrate theoretically and experimentally that our proposed **MLDSP-MA** has better performance and competitiveness.

2. RELATED WORK

As mentioned above, conversational sentiment prediction for intelligent chatbots has been a research hotspot in recent years. However, the related work supporting this research involves many fields.

2.1 Sentiment Extraction

Since dialogue sentiment prediction originates from sentiment extraction, it is an important foundation for dialogue sentiment prediction. Guo et al. (2021) proposed an INIT-CNN model to extract sentiments from Weibo, which employs internet slangs, negative words, and emotional symbols to assist in sentiment extraction. Nezhad et al. (2022) proposed a CNN-LSTM model to extract sentiments from Twitter, considering the relationship between attitude and sentiment. Wang et al. (2022) proposed a model called HASA to extract review sentiments on commodities from e-commerce websites, which has certain reference value in recommending products. Similarly, for the commodity reviews on e-commerce websites, Dadhich et al. (2022) proposed a hybrid rule-based model, while Zhang et al. (2022) proposed a model based on PMI and DC-PNC, which uses sequence labeling and syntactic analysis methods to extract new emotional words from commodity reviews. Bhuvaneshwari et al. (2022) employed self-attention mechanism to extract sentiments from commodity review information and solved the problem of data sparsity. Li et al. (2022) proposed a method called BiERU to extract sentiments from dialogue, which can quickly extract sentimental information. Qin et al. (2021) proposed a Co-GAT model based on cooperative interaction graph attention network for dialogue behavior recognition and sentiment classification. Chen et al. (2022) studied the transfer of sentiments between individual speakers and proposed a transformer-based dialogue sentiment extraction model. Ji et al. (2022) proposed an AFR-BERT model for dialogue sentiment extraction based on attention mechanisms, considering the effective fusion and correlation of multimodal data. In summary, the rapid development of sentiment extraction has laid a solid foundation for dialogue sentiment prediction.

2.2 Sentiment Analysis

Sentiment analysis is a fine-grained task in the field of natural language processing. It is the core technology and challenge of dialogue sentiment prediction. Based on customer reviews on the Amazon shopping platform, Vanaja et al. (2018) analyzed the sentimental characteristics of the reviews and provided a rating method. Wang et al. (2018) proposed an LSTM model based on segment attention, which effectively analyzes the dependency between entities and sentimental expressions through linear-chain conditional random field (CRF). Huang et al. (2022) proposed an aspect-based sentiment analysis model, which calculates sentiment scores by designing two asymmetrical context position weighting functions. Chen et al. (2019) developed a Transfer Capsule Network (TransCap) model that integrates sentence-level semantic analysis, aspect-level semantic analysis, and document-level semantic analysis. To analyze the importance of sentences in documents, Choi et al. (2020) proposed a deep neural network-based sentence classification model that automatically calculates the importance of sentences in documents through gate mechanisms. Similarly, Cao et al. (2022) also proposed a method to analyze the importance of sentences using user-enhanced pre-trained language models and user identity information. Chen et al. (2021) considered users' commenting and proposed the HUSN model for analyzing the importance of sentences. Obviously, sentiment analysis not only involves complex neural network models but also relates to human behavior and psychology. Therefore, it is the core technology and challenge of research on dialogue sentiment prediction.

2.3 Sentiment Estimation

Sentiment estimation is the process of predicting the sentiment of a sentence before it is spoken. Sentiment estimation is a practical application of dialogue sentiment prediction and has various applications in intelligent chatbots, online translation, and online diagnosis. For example, Gaonkar et al. (2020) developed a sentiment tag prediction system that can predict the sentiments of characters in stories. Li et al. (2019) established a unified model for sentiment extraction, sentiment analysis, and sentiment estimation, using joint modeling and unified labeling methods. Piao et al. (2018) built a financial text sentiment estimation system that combines RNN, CNN, and ridge regression methods. Wang et al. (2020) proposed a neural simulation model that predicts dialogue sentiments using pair-wise and seq-wise relationships. Bothe et al. (2017) proposed a dialogue sentiment prediction model based on recursive neural networks that leverages the contextual relationships of utterances. Zou et al. (2022) discussed the distinction between emotion and sentiment and used terms such as joy, surprise, trust, anticipation, anger, disgust, fear, and sadness in dialogues to assist in dialogue sentiment prediction. Cai et al. (2020) developed a dialogue recommendation system that predicts user intent and provides relevant recommendations. These studies have made dialogue sentiment prediction a hotspot in natural language applications. However, sentiment estimation still faces challenges in dealing with long dialogues with multiple topics.

3. MULTI-ROUND LONG DIALOGUE SENTIMENT PREDICTION BASED ON MULTIDIMENSIONAL ATTENTION

Problem desc.: Let $\mathrm{D} = \{U_A^1, U_B^1; \dots; U_A^t, U_B^t; \dots; U_A^n, U_B^n\}$ represent a multi-round conversation, where U_A^t and U_B^t represent the utterances of participant A and B respectively in the *t*-th round of conversation. Taking U_A^t as an example, each utterance consists of *m* words, i.e., $U_A^t = \{w_{A,1}^t, w_{A,2}^t, \dots, w_{A,m}^t\}$, where $w_{A,m}^t$ represents the *m*-th word in this utterance. In order to predict sentiment from the conversation, we introduce symbols S_A^t and S_B^t, which respectively represent the sentiment of A and B in the *t*-th round of conversation, as shown in Equation (1).

$$S_A^t \sim P(D_A^t),\ S_B^t \sim P(D_B^t), \tag{1}$$

where $D_A^t = \{U_A^1, U_B^1; \dots; U_A^{t-1}, U_B^{t-1}\}$ represents the previous *t*-1 rounds of conversation for A, and *P* represents the probability prediction. Obviously, the sentiment S_A^t of A in the *t*-th round of conversation need to be predicted based on D_A^t. Similarly, $D_B^t = \{U_A^1, U_B^1; \dots; U_A^{t-1}, U_B^{t-1}, U_A^t\}$ represents the previous *t*-1 rounds of conversation for B. It should be noted that U_A^t is also included in D_B^t. Therefore, the task of sentiment prediction in multi-round dialogues is to extract sentiment information from the given historical multi-round conversations, and then, to predict the sentiment of the participants in the next moment.

As mentioned earlier, as the number of conversation rounds increases, the topics discussed by both parties and the corresponding sentiments change. However, existing methods do not take this characteristic into account. In light of this situation, we propose a multi-dimensional attention-based model for sentiment prediction in multi-round dialogues called **MLPDSP-MA**, which aims to address the problem of too many historical conversation topics affecting the accuracy and efficiency of sentiment prediction. **MLPDSP-MA** consists of the DPI extraction module, LAF module, and GSA module, where DPI refers to Dependency-Persistence-Infectivity, LAF refers to Local Attention Fusion, and GSA refers to Global Sentiment Attention, as shown in Figure 2. In **MLPDSP-MA**, to easily capture different topics and their corresponding sentiments, we introduce the sliding window technique. Assuming the size of the sliding window is *m*, as the sliding window D_m moves, the dialogue within the window is treated as one topic, and the sentiment of the next utterance within the window is regarded as the prediction target. The sliding window can capture the logical relationships in the sequence data, so it is leveraged by **MLPDSP-MA** to predict changes in sentiment, as shown in Equation (2).

$$S^t \sim P(D_m), \tag{2}$$

where S^t represents the sentiment in the *t*-th round of conversation, and *P* is the probability function.

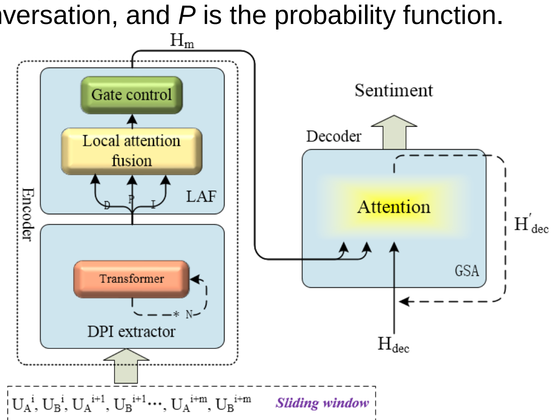

Figure 2: Illustration of the proposed **MLDSP-MA** framework.

In Figure 2, $U_A^i, U_B^i; U_A^{i+1}, U_B^{i+1}; \dots; U_A^{i+m}, U_B^{i+m}$ represent the dialogues collected by the sliding window. The DPI module is used to extract contextual dependencies, sentiment persistence, and sentiment infectivity from the collected dialogue. The LAF module is used to perform attention-based fusion of these contextual dependencies, sentiment persistence, and sentiment infectivity. The gate control represents a sigmoid-based gating mechanism. H_m is the fused sentiment information for the current dialogue. The GSA module is exploited to learn the sentiment in the dialogue from a global perspective and predict future sentiment. The connections between these modules are shown in Figure 2. From Figure 2, we can see that the DPI extractor and LAF form the Encoder, while the GSA forms the Decoder.

3.1 DPI Extraction and Local Attention Fusion

As mentioned abover, DPI extraction and local attention fusion are integral parts of **MLPDSP-MA**. The former is used to extract dependencies, persistence, and infectivity in conversations, while the latter is used to fuse these features. DPI extraction is the foundation of **MLPDSP-MA** because in multi-round dialogues, contextual dependencies, sentimental persistence, and sentimental infectivity run through the dialogues, and they play the role of reflecting the topics, embodying sentiments, and anticipating changes, as shown in Table 1.

Features	Explanations	Examples	rules
Dependency(D)	Also known as context dependency. It comes from the similarity and continuity of topics in adjacent dialogues, so there are dependencies between dialogues.	If the conversation is about the movie, then the sentiment revolves around the movie. If the conversation is about football, then the sentiment revolves around football. Therefore, sentiments depend on context.	■ Different topics have different emotions. ■ Same topics have sentimental impact. ■ Topics with large time spans have less mutual sentimental impact.
Persistency(P)	It refers to the basically unchanged nature of sentiments in the same topic. This is because the sentiments of the participants in the conversation about a certain topic always run through the topic.	If the dialogue participant shows a negative attitude towards FC Barcelona, he will convey throughout the conversation.	■ If the topic does not change, we can infer the sentiments behind the conversation participant based on his previous sentiments.
Infectivity(I)	It means that the sentiments of one of the participants in the conversation affect the other. This is because the dialogue process is a process of empathy.	Dialogue participant B said that "whether FC Barcelona can keep the Champions League is a question", which led to dialogue participant A saying that Barca were "screwing up ".	■ The sentiments of one participant in the conversation are known, and if the topic has not changed, the sentiments of the other party of the participant can be speculated.

Table 1: Role of DPI in Multi-round Dialogues

Table 1 shows the roles, examples, and patterns of the contextual dependencies, sentimental persistence, and sentimental infectivity. These features are acquired by the DPI extraction module, which exploits Transformers to extract from the current dialogue (within a sliding window), as shown in equations (3), (4), and (5).

$$CD = \text{Transformer}^D(D_m), \quad (3)$$

$$SP = \text{Transformer}^P(D_m), \quad (4)$$

$$SI = \text{Transformer}^I(D_m), \quad (5)$$

In those equations, CD represents contextual dependencies, SP represents sentimental persistence, and SI represents sentimental infectivity. It is worth noting that CD, SP, and SI extraction do not share the same Transformer, which effectively improves the model's response speed, as shown in Figure 3.

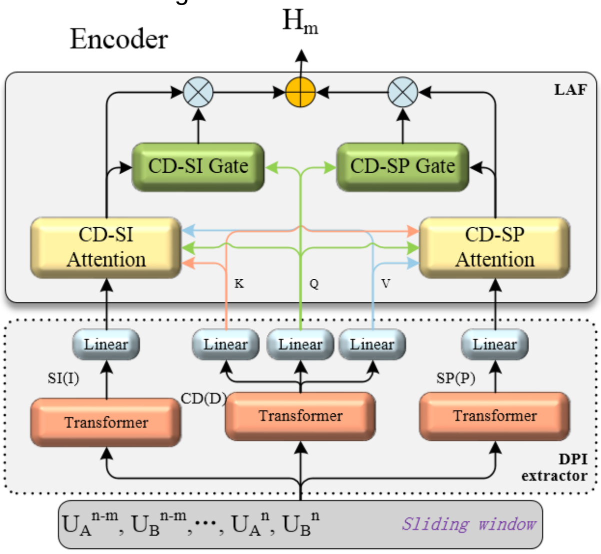

Figure 3: Encoder of MLDSP-MA. The encoder of **MLDSP-MA** includes LAF and DPI extractors, and its input is $U_A^{n-m}, U_B^{n-m}; \dots; U_A^n, U_B^n$.

In Figure 3, the DPI extractor is located below, while the LAF is located above. For the DPI extractor, it consists of three Transformers to extract SI(I), CD(D), and SP(P) respectively. For the LAF, it fuses the contextual dependencies and sentimental infectivity of the conversation to obtain the local attention fusion information of CD-SI. Then, it fuses the contextual dependencies and sentimental persistency of the conversation to obtain the local attention fusion information of CD-SP. The calculation methods for CD-SI and CD-SP are shown in Equation (5) and Equation (6).

$$\mathrm{CD-SI} = \mathrm{LAF}(\mathrm{CD}, \mathrm{SI}), \tag{5}$$

$$\mathrm{CD-SP} = \mathrm{LAF}(\mathrm{CD}, \mathrm{SP}), \tag{6}$$

In Equation (5) and (6), LAF adopts an attention mechanism, and exploits the contextual dependencies of the dialogue as query (Q), key (K), and value (V), as shown in Equation (7).

$$\mathrm{CD}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) = Att(D_m;\ W_Q, W_K, W_V), \tag{7}$$

where *Att* represents the self-attention mechanism of the Transformer, and W_Q, W_K, W_V are learnable parameters.

For CD-SI, since the interaction between CD and SI is a learning process, the optimal interaction values can be dynamically captured by setting learning parameters. For example, λ_{KI} and λ_{VI} are set as learnable parameters for $\widehat{K}_I$ and $\widehat{V}_I$, as shown in Equation (8).

$$\begin{bmatrix}\widehat{K}_I\\ \widehat{V}_I\end{bmatrix} = (1\text{-}\begin{bmatrix}\lambda_{KI}\\ \lambda_{VI}\end{bmatrix})\mathrm{CD}(\mathrm{K}, \mathrm{V}) + \begin{bmatrix}\lambda_{KI}\\ \lambda_{VI}\end{bmatrix}(SI\begin{bmatrix}P_{KI}\\ P_{VI}\end{bmatrix}), \tag{8}$$

where $\widehat{K}_I$ and $\widehat{V}_I$ represent the key and value vectors of CD-SI, and P_{KI} and P_{VI} are learnable matrices. To ensure good learnability of Equation (8), LAF exploit a gating mechanism based on sigmoid to calculate λ_{KI} and λ_{VI}, as shown in Equation (9).

$$\begin{bmatrix}\lambda_{KI}\\ \lambda_{VI}\end{bmatrix} = \sigma(\mathrm{CD}(\mathrm{K}, \mathrm{V})\begin{bmatrix}W_{KI1}\\ W_{VI1}\end{bmatrix} + SI\begin{bmatrix}P_{KI}\\ P_{VI}\end{bmatrix}\begin{bmatrix}W_{KI2}\\ W_{VI2}\end{bmatrix}), \tag{9}$$

where W_{KI1}, W_{VI1}, W_{KI2}, and W_{VI2} are all learnable parameters. The output of CD-SI is obtained by scaled dot-product and denoted as $H_{\mathrm{CD-SI}}$, as shown in Equation (10).

$$H_{\mathrm{CD-SI}} = Softmax\left(\frac{Q\widehat{K}_I^T}{\sqrt{d_k}}\right)\widehat{V}_I, \tag{10}$$

where *T* denotes transpose, d_k represents the dimensionality of vectors, and $\widehat{K}_I$ and $\widehat{V}_I$ are calculated by Equation (8).

Similarly, for CD-SP, its specific calculation equations are shown in (11), (12), and (13).

$$\begin{bmatrix}\widehat{K}_P\\ \widehat{V}_P\end{bmatrix} = (1\text{-}\begin{bmatrix}\lambda_{KP}\\ \lambda_{VP}\end{bmatrix})\mathrm{CD}(\mathrm{K}, \mathrm{V}) + \begin{bmatrix}\lambda_{KP}\\ \lambda_{VP}\end{bmatrix}(SP\begin{bmatrix}U_{KP}\\ U_{VP}\end{bmatrix}), \tag{11}$$

$$\begin{bmatrix}\lambda_{KP}\\ \lambda_{VP}\end{bmatrix} = \sigma(\mathrm{CD}(\mathrm{K}, \mathrm{V})\begin{bmatrix}W_{KP1}\\ W_{VP1}\end{bmatrix} + SP\begin{bmatrix}U_{KP}\\ U_{VP}\end{bmatrix}\begin{bmatrix}W_{KP2}\\ W_{VP2}\end{bmatrix}), \tag{12}$$

$$H_{\mathrm{CD-SP}} = Softmax\left(\frac{Q\widehat{K}_P^T}{\sqrt{d_k}}\right)\widehat{V}_P, \tag{13}$$

where $\widehat{K}_P$ and $\widehat{V}_P$ represent the key and value vectors of CD-SP, λ_{KP}, λ_{VP}, W_{KP1}, W_{VP1}, W_{KP2}, and W_{VP2} are all learnable parameters, and U_{KP} and U_{VP} are learnable matrices. $H_{\mathrm{CD-SP}}$ is the output of CD-SP, which adopts the scaled dot-product calculation method.

Since CD-SI and CD-SP have different impacts on the subsequent dialogues' sentiments, LAF uses a weighted sum method to fuse $H_{\mathrm{CD-SI}}$ and $H_{\mathrm{CD-SP}}$, as shown in Equation (14), (15), and (16).

$$g_{CD-SI} = (Q \oplus H_{CD-SI})W_C + b_C, \tag{14}$$

$$g_{CD-SP} = (Q \oplus H_{CD-SP})W_S + b_S, \tag{15}$$

$$H_m = g_{CD-SI} \otimes H_{CD-SI} + g_{CD-SP} \otimes H_{CD-SP}, \tag{16}$$

where W_C and b_C are learnable parameters for CD-SI, W_S and b_S are learnable parameters for CD-SP, $\oplus$ represents concatenation, and *Q* represents Query. H_m is the sentimentally integrated information of the current dialogue.

3.2 Global Sentiment Fusion

As mentioned above, historical dialogues are intercepted by the sliding window and fully traversed as the window moves. In this process, the contextual dependency, sentimental persistency, and sentimental infectivity of the current dialogue are respectively learned and integrated into H_m. However, this is a local learning process that lacks learning of comprehensive sentiments. To this end, we propose the Global Sentiment Attention (GSA) module, which attempts to learn sentiments in the dialogue from a global perspective. As shown in Figure 4, that is essentially a decoder.

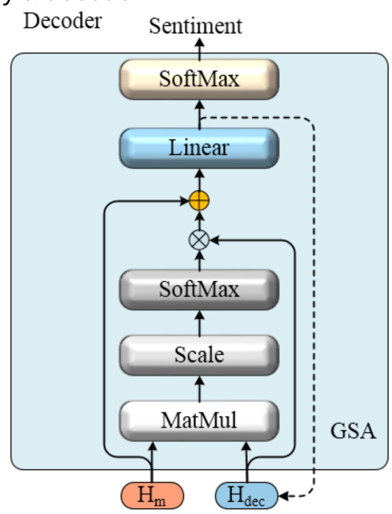

Figure 4: Global sentiment fusion module GSA. GSA is the decoder of MLDSP-MA. H_m is the input of GSA and also the output of the aforementioned encoder, and H_{dec} is the sentimental fusion information outputted by the decoder.

In Figure 4, H_m is the output of the aforementioned encoder, which serves as the input to GSA, and H_{dec} is the sentiment fusion information iteratively outputted by the decoder. It is worth noting that the initial value of H_{dec} is null, and after each sentiment prediction, its value is updated once.

Obviously, GSA is also an information fusion module that integrates the outputs of the encoder and decoder, as shown in Equation (17).

$$H, H'_{dec} = GSF(H_m, H_{dec}), \tag{17}$$

where *H* is the sentiment prediction, H'_{dec} is the updated sentiment fusion information from the decoder. Like the LAF module, GSA also introduces a scaled dot product attention mechanism, as shown in Equation (18).

$$\widehat{\boldsymbol{H}}_{\boldsymbol{dec}} = Softmax\left(\frac{H_m H_{dec}^T}{\sqrt{d_k}}\right) H_{dec}, \tag{18}$$

Since H_{dec} is corresponded to H_m , H'_{dec} overall reflects the contextual dependency, sentimental persistency, and sentimental infectivity of multi-round dialogues. To strengthen global sentiments, GSA employs a residual connection mechanism, as shown in Equation (19).

$$\widehat{H} = H_m + \widehat{H}_{dec}, \tag{19}$$

To fully perform sentiment prediction and extract sentiment fusion information, after obtaining $\widehat{H}$, the decoder further processes $\widehat{H}$ with a fully connected layer, as shown in Equation (20).

$$H, H'_{dec} = Linear(\widehat{H}), \tag{20}$$

Finally, through activation function processing, the ultimate predicted sentiment is obtained.

3.3 Loss Functions

Multi-round long dialogue sentiment prediction employs cross-entropy loss function to implement iterative training. Assuming there are *m* dialogues, the corresponding loss function $\mathcal{L}$ is shown in Equation (22).

$$\mathcal{L} = -\frac{1}{m}\sum_{i=1}^{m}[p_{senti}^{i} \, log(\hat{p}_{senti}^{i}) + (1 - p_{senti}^{i}) \, log(1 - \hat{p}_{senti}^{i})], \tag{22}$$

where p_{senti} is the true sentiment of the *i*-th dialogue, while $\hat{p}_{senti}^{i}$ represents the predicted sentiment of the *i*-th dialogue.

4. EXPERIMENTS

To validate the effectiveness of our proposed **MLDSP-MA**, we conducted comparative experiments with relevant baseline models. Performance experiments and ablation studies were performed, similar to the state-of-the-art methods (Wang et al., 2020; Zou et al., 2022).

4.1 Datasets

To validate the effectiveness of **MLDSP-MA**, we utilized the KdConv dataset and the IEMOCAP dataset, where the former is a Chinese dialogue dataset and the latter is an English dialogue dataset, as follows:

(1) KdConv dataset (Zhou et al., 2020): This is a Chinese multi-round dialogue dataset consisting of 4,500 dialogue records, which can be used for predicting dialogue sentiments. The dialogues in this dataset cover topics such as movies, music, and travel. The average length (number of sentences/utterances) of each record is 20, with 90% of the records having lengths between 14 and 24. It is important to note that the sentiment distribution in this dataset is as follows: neutral 50.44%, positive 44.12%, negative 5.44%. Obviously, this imbalanced distribution poses a significant challenge for the sentiment prediction task.

(2) IEMOCAP dataset (Ghosal et al., 2020): This is a multimodal English dialogue dataset that contains 152 long dialogue records. One notable characteristic of this dataset is that each record contains a large number of sentences, with an average of 60 sentences, making it a typical long dialogue dataset. The sentiment distribution in this dataset is as follows: neutral 62.85%, negative 22.51%, positive 14.64%.

Both datasets mentioned above are divided into three parts: training set, validation set, and testing set. The proportions of these three parts are 8:1:1, where 80% of the data is used for model training, 10% is used for adjusting model the hyper-parameters of the model and preventing overfitting, and 10% is used for testing the model.

4.2 Evaluation Indicators

To evaluate the experimental effects, we adopted evaluation metrics commonly used in multi-round dialog research (Vinod and Sheeja, 2023; Wu et al., 2023), namely accuracy and Micro-F1 score. Micro-F1 involves precision, recall, and *F1* score, as shown in Equation (23), (24), (25), (26), and (27).

$$\text{Acc} = \frac{TP+TN}{TP+FP+TN+FN}, \tag{23}$$

$$P = \frac{TP}{TP+FP}, \tag{24}$$

$$R = \frac{TP}{TP+FN}, \tag{25}$$

$$F1 = \frac{2*\text{P}*\text{R}}{\text{P+R}}, \tag{26}$$

$$\text{Macro} - \text{F1} = \frac{1}{C}\sum_{i=1}^{C} F_i, \tag{27}$$

To eliminate experimental result biases caused by dataset partitioning, we employed the 10-fold Cross Validation method for verification. The average of all experimental results was considered as the final experimental result.

4.3 Baselines and Implementations

Due to the reference and comparative roles of baseline methods for our proposed **MLPSP-MA**, we selected 8 typical baseline methods for dialogue sentiment prediction, namely RNN (Khanpour et al., 2016), LSTM (Poria et al., 2017), DialogueRNN (He et al., 2016), BERT (Devlin et al., 2019), DialogueGCN (Ghosal et al., 2019), NSF (Wang et al., 2020), STEM (Zhen et al., 2023), and EASF (Zou et al., 2022). These baseline methods are introduced as follows:

(1) RNN (Khanpour et al., 2016): It is essentially a dialogue sentiment prediction method based on RNN, utilizing the neural network's memory function for past states, making it a preferred method for handling sequential data.

(2) LSTM (Poria et al., 2017): It is essentially a dialogue sentiment prediction method based on LSTM, overcoming the problem of gradient vanishing in RNN methods, hence it has been widely adopted in recent years.

(3) DialogueRNN (He et al., 2016): It is an improved RNN model that uses three Gate Recurrent Units to capture sentimental information in dialogues.

(4) BERT (Devlin et al., 2019): It is a bidirectional encoder that utilizes the masked language model method. This approach leverages the pre-training results of BERT, allowing it to achieve better performance.

(5) DialogueGCN (Ghosal et al., 2019): It is a dialogue sentiment prediction model based on GCN, considering the relationships between dialogue participants, and also introducing graph attention mechanisms.

(6) NSF (Wang et al., 2020): It is a context-based dialogue sentiment prediction model that considers the influence of utterance order on sentiments.

(7) STEM (Zhen et al., 2023): It is a feature extraction-based dialogue sentiment prediction model that employs data mining methods to discover latent features and considers the relationship between sentimental and behavioral aspects of dialogue participants.

(8) EASF (Zou et al., 2022): It is an improved NSF model that considers the distinction between sentiment and emotion, using emotion information to assist in dialogue sentiment prediction.

In this study, pre-trained model BERT-Base was used, with parameters set to 12-heads, 12-layers, and 768-hidden. The KdConv dataset and IEMOCAP dataset were leveraged for model training. When using the KdConv dataset, the batch size was set to 50, the number of epochs was set to 100, the optimizer was set to Adam, the initial learning rate was set to 0.01, and the learning rate was multiplied by 0.5 every 10 epochs for decay. Similarly, when using the IEMOCAP dataset, the batch size was set to 5, the number of epochs was set to 50, the optimizer was set to Adam, the initial learning rate was set to 0.01, and the learning rate was reduced to half every 10 epochs. During the training process, if the model's performance improved, it was saved; otherwise, it was not saved.

Additionally, to conduct experimental research, we implemented all models using the PyTorch framework. We also performed experiments to investigate the impact of the sliding window size. The sliding window size was gradually increased from 1 until the optimal prediction result and stable evaluation metrics were achieved.

4.4 Performance Study

To demonstrate the competitiveness of our proposed **MLPSP-MA**, we conducted experimental comparisons with baseline models RNN, LSTM, DialogueRNN, BERT, DialogueGCN, NSF, STEM, and EASF, as shown in Table 2 and Table 3. In the experiments, we evaluated the competitiveness of the models from different perspectives of metrics and datasets. The metrics used were Accuracy and Micro-F1 score, and the datasets used were KdConv and IEMOCAP. Furthermore, to indicate models with better performance, we bold them and underline the top three models.

Model	Accurate (%)	
	KdConv	IEMOCAP
RNN[2016]	38.54	30.20
LSTM[2017]	39.61	32.71
DialogueRNN[2016]	41.14	34.89
BERT[2019]	42.31	36.87
DialogueGCN[2019]	43.38	37.46
NSF[2020]	44.07	40.34
STEM[2023]	43.48	40.34
EASF[2022]	44.95	42.22
MLDSP-MA (ours)	**46.77**	**43.41**

Table 2: Performance comparison of Accuracy on the datasets KdConv and IEMOCAP

From Table 2, it can be seen that compared to baseline methods, our **MLPSP-MA** achieves the highest accuracy in both Chinese dialogue dataset KdConv and English dialogue dataset IEMOCAP. Specifically, on the KdConv dataset, **MLPSP-MA** achieves an accuracy of 46.77%, whereas RNN and LSTM are below 40%, DialogueRNN, BERT, DialogueGCN, NSF, and EASF are below 45%. In terms of performance improvement, **MLPSP-MA** outperforms RNN by 8.23% and outperforms state-of-the-art EASF by 1.82%. Similarly, on the IEMOCAP dataset, **MLPSP-MA** achieves an accuracy of 43.41%, while RNN, LSTM, DialogueRNN, BERT, and DialogueGCN are below 40%, and NSF and EASF are below 43%. In terms of performance improvement, **MLPSP-MA** outperforms RNN by 13.21% and outperforms EASF by 1.19%.

These data validate that our **MLPSP-MA** is a competitive model that utilizes attention mechanisms to identify relevant sentences and words related to sentiment prediction. The attention mechanism of **MLPSP-MA** is a multi-dimensional attention that separately identifies sentences related to contextual dependence, sentiment persistence, and sentiment infectivity from the historical dialogues, and learns local and global attentions for them. This is also the reason why **MLPSP-MA** outperforms the baselines. We believe that in multi-round long dialogues, if attention mechanisms are not utilized or if the attention mechanisms are not targeted enough, the large number of dialogues and frequent changes in sentiment will severely affect the accuracy of sentiment prediction.

Model	Micro-F1 (%)	
	KdConv	IEMOCAP
RNN[2016]	32.24	25.68
LSTM[2017]	33.04	28.28
DialogueRNN[2016]	34.37	30.59
BERT[2019]	35.95	32.28
DialogueGCN[2019]	36.96	32.82
NSF[2020]	37.81	35.75
STEM[2023]	35.98	32.66
EASF[2022]	38.83	37.57
MLDSP-MA (ours)	**41.55**	**39.24**

Table 3: Performance comparison of Micro-F1on the datasets KdConv and IEMOCAP

From Table 3, it can be observed that in terms of Micro-F1 performance, **MLPSP-MA** outperforms all the baselines. On the KdConv dataset, it achieves 41.55%, while RNN, LSTM, DialogueRNN, and BERT are all below 36%, and DialogueGCN, NSF, and EASF are all below 39%. In terms of performance improvement, **MLPSP-MA** improves upon RNN by 9.31% and EASF by 2.72%. Similar results are observed on the IEMOCAP dataset. That is, our

MLPSP-MA achieves 39.24%, while RNN, LSTM, DialogueRNN, BERT, and DialogueGCN are all below 33%, and NSF and EASF are all below 37.6%. In terms of performance improvement, **MLPSP-MA** outperforms RNN by 13.56% and EASF by 1.67%. Therefore, from the perspective of the Micro-F1 metric, the experiments indicate that **MLPSP-MA** has reached an advanced level.

In summary, from different perspectives of datasets and evaluation metrics, **MLPSP-MA** has been proven to be a competitive model.

4.5 Ablation Study

To verify the effectiveness of the key components of **MLPSP-MA**, ablation studies need to be conducted. We performed ablation experiments on the KdConv dataset and IEMOCAP dataset, with the evaluation metrics being Accuracy and Micro-F1 score. In the ablation experiments, "w/o. DPI_LAF" indicates that the window-based DPI and LAF modules are removed, and "w/o. GSA" indicates that the GSA module is removed. The results of the ablative experiments are shown in Table 4.

Model	KdConv		IEMOCAP	
	Accuracy	Micro-F1	Accuracy	Micro-F1
w/o. DPI_LAF	42.02	37.48	41.17	36.41
w/o. GSA	45.35	40.43	43.08	37.76
MLDSP-MA (ours)	**46.77**	**41.55**	**43.41**	**39.24**

Table 4: Results of ablation study

Analysis: From Table 4, it can be seen that when removing DPI_LAF, i.e., "w/o. DPI_LAF", the accuracy performance of the model decreases by 4.75% and the Micro-F1 performance decreases by 3.08% on the KdConv dataset. This is because the extraction of contextual dependence, sentiment persistence, and sentiment infectivity is crucial for predicting dialogue sentiment, and it significantly improves the performance of dialogue sentiment prediction. If the GSA module is removed, i.e., "w/o. GSA", the accuracy performance of the model decreases by 1.42% and the Micro-F1 performance decreases by 1.12% on the KdConv dataset. This indicates that global sentiment attention can further improve the performance of dialogue sentiment prediction. Similarly, the same results can be obtained on the IEMOCAP dataset. It is worth noting that removing DPI_LAF leads to a more significant performance decrease compared to removing GSA. This suggests that the window-based DPI and LAF proposed in this paper are important improvements in the field of dialogue sentiment prediction.

As mentioned earlier, the **MDSP-MA** model, which is a multi-round long dialogue sentiment prediction model based on multi-dimensional attention, introduced the sliding windows. Through the sliding windows, the negative impact of earlier topics and sentiments on the current dialogue sentiment is weakened. This is because in multi-round long dialogues, the current dialogue topic has already diverged from earlier topics, and the dialogue sentiment has also changed multiple times. Therefore, in machine learning, selecting too many historical dialogues can sometimes interfere with the extraction of current dialogue sentiment. The **MDSP-MA** model introduces a sliding window of size *m*, where *m* is a hyper-parameter. In order to analyze the influence of different values of *m* on experimental results and observe the optimal value of *m* for different datasets, corresponding experiments were conducted, as shown in Table 5.

m	KdConv		IEMOCAP	
	Accuracy	Micro-F1	Accuracy	Micro-F1
1	39.95	34.49	33.99	29.69
2	41.67	36.21	36.47	32.29
3	43.02	37.56	38.55	34.48
4	44.10	38.64	40.04	36.01
5	44.84	39.49	41.43	37.40
6	45.35	40.08	42.52	38.43
7	45.88	40.66	**43.41**	**39.24**
8	46.37	41.16	43.31	39.13
9	46.72	41.47	43.31	39.06
10	**46.77**	**41.55**	43.11	38.91
11	46.73	41.45	43.31	39.18

Table 5: Experiments on the influence of *m* values

Analysis: From Table 5, it can be seen that accuracy and Micro-F1 vary with different values of *m*. However, when *m*=10, the **MDSP-MA** model achieves the best performance with accuracy of 46.77 and Micro-F1 of 41.55 on the KdConv dataset. This indicates that larger *m* is not necessarily better, nor is smaller *m*. Similarly, when *m*=7, the **MDSP-MA** model achieves the best performance with accuracy of 43.41 and Micro-F1 of 39.24 on the IEMOCAP dataset. Therefore, the optimal value of *m* varies for different datasets, and for the same dataset, there exists an optimal *m* value.

5. Conclusion

This paper has demonstrated our proposed **MDSP-MA** model, which introduces the sliding window to capture the topics of dialogues and incorporates a multi-dimensional attention mechanism to predict the sentiment of dialogues. Due to the existing methods' low accuracy in handling long dialogue data, we were motivated to research topic-based dialogue sentiment prediction methods and proposed the **MDSP-MA** model. In order to capture different topics, we exploited the sliding window technique, which not only traverses all historical dialogues but also focuses on the dialogue within the current topic. To improve the accuracy of dialogue sentiment prediction, we treated the pairwise interactions of contextual dependency, sentiment persistency, and sentiment infectivity as local attention, and treated the comprehensive sentiment iteratively trained from historical dialogue data as global attention. Finally, the two types of attention are merged. Through sufficient experiments on two publicly available datasets, the experimental results show that our model surpasses current state-of-the-art models. Our future work will explore adaptive sliding window techniques (Su et al., 2023; Feng et al., 2023), which can automatically adapt to the length of different topics. We believe this will inspire meaningful approaches.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical Approval

Not applicable.

Competing interests

Authors declare that *there are no financial and personal relationships with other people or organizations that can inappropriately influence the work*, *there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled*.

6. Bibliographical References

Bhuvaneshwari, P., Rao, A.N., Robinson, Y.H., et al.(2022). Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model. *Multimedia Tools and Applications*, 81(9):12405-12419.

Bothe, C., Magg, S., Weber, C., et al. (2017). Dialogue-based neural learning to estimate the sentiment of a next upcoming utterance. In *Proceedings of 26th International Conference on Artificial Neural Networks*, Alghero, Italy, September 11-14, pages477-485.

Cai, W., Chen, L. (2020). Predicting user intents and satisfaction with dialogue-based conversational recommendations, In *Proceedings of the 28th ACM Conference on User Modeling*, Adaptation and Personalization, pages33-42.

Calabrese, A., Rivoli, A., Sciarrone, F., Temperini, M. (2023). An intelligent chatbot supporting students in massive open online courses. In *Proceedings of 21st International Conference on Web-Based Learning (ICWL 2022)*, vol. 13869 LNCS, pages 190-201.

Cao, X., Yu, J., Zhuang, Y. (2022). Injecting user identity into pretrained language models for document-level sentiment classification. *IEEE Access*, 10: 30157-30167.

Chen, C., Zhou, X., Wang, Z., Zhang, M. (2022). Dialog sentiment analysis with multi-party attention. *Journal of Chinese Information Processing*, 36(12):173-181.

Chen, J., Yu, J., Zhao, S., et al. (2021). User's review habits enhanced hierarchical neural network for document-level sentiment classification. *Neural Processing Letters*, 53(3):2095-2111.

Chen, Z. and Qian, T. (2019). Transfer capsule network for aspect level sentiment classification. In Proceedings of the 57th annual meeting of the association for computational linguistics, pages 547-556.

Choi, G., Oh, S., Kim, H. (2020). Improving document-level sentiment classification using importance of sentences. Entropy, 22(12):1-11.

Dadhich, A., Thankachan, B. (2022). Sentiment analysis of amazon product reviews using hybrid rule-based approach. In *Proceedings of Smart Systems: Innovations in Computing (SSIC 2021)*, Springer Singapore, pages 173-193.

Devlin, J., Chang, M.W., Lee, K., et al. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. 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)*, pages4171-4186.

Feng, B., Ding, Z., Jiang, C. (2023). FAST: A forecasting model with adaptive sliding window and time locality integration for dynamic cloud workloads. *IEEE Transactions on Services Computing*, 16(2): 1184-1197.

Ghosal, D., Majumder, N., Gelbukh, A., et al. (2020). COSMIC: commonSense knowledge for emotion identification in conversations. In *Proceedings of the Association for Computational Linguistics (EMNLP 2020)*, pages 2470-2481.

Ghosal, D., Majumder, N., Poria, S., et al. (2019). 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 2019)*, pages154-164.

Guo, X., Lai, H., Yu, Z., Gao, S., Xiang, Y. (2021). Emotion classification of case-related microblog comments integrating emotional knowledge. *Chinese Journal of Computers*, 44(3):564-578.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pages 770-778, December 9.

Huang, B., Guo, R., Zhu, Y., et al. (2022). Aspect-level sentiment analysis with aspect-specific context position information. *Knowledge-Based Systems*, 243:1-11.

Huang, J. (2023). Research on chat robot based on seq2seq model. In *Proceedings of Sixth International Conference on Intelligent Computing*, Communication, and Devices, ICCD 2023, pages1-8.

Ji, M., Zhou, J., Wei, N. (2022). AFR-BERT: Attention-based mechanism feature relevance fusion multimodal sentiment analysis model. *Plos one*, 17(9):1-20.

Khanpour, H., Guntakandla, N., Nielsen, R. (2016). Dialogue act classification in domain-independent conversations using a deep recurrent neural network. In *Proceedings of the 26th international conference on computational linguistics: Technical papers*, pages 2012-2021.

Li, W., Shao, W., Ji, S., et al. (2022). BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing*, 467:73-82.

Li, X., Bing, L., Li, P., et al. (2019). A unified model for opinion target extraction and target sentiment prediction, In *Proceedings of the AAAI conference on artificial intelligence*, 33(1): 6714-6721.

Nezhad, Z.B. and Deihimi, M.A. (2022). Twitter sentiment analysis from Iran about COVID 19 vaccine. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 16(1):1-14.

Piao, G. and Breslin, J.G. (2018). Financial aspect and sentiment predictions with deep neural networks: an ensemble approach. In *Proceedings of the The Web Conference 2018*, pages1973-1977.

Poria, S., Cambria, E., Hazarika, D., et al. (2017). Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 873-883.

Qin, L., Li, Z., Che, W., et al. (2021). Co-gat: A co-interactive graph attention network for joint dialog act recognition and sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages13709-13717.

Su, L., Zhang, C., Tan, H., Qin, H., Wang, S., Wu, G., Xue, Y. (2023). Research on electricity load demand forecasting algorithm based on adaptive sliding window method and attention mechanism, In *Proceedings of 4th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT 2023)*, pages 456-460.

Vanaja, S. and Belwal, M. (2018). Aspect-level sentiment analysis on e-commerce data. In *Proceedings of 2018 International conference on inventive research in computing applications (ICIRCA 2018)*, IEEE, pages1275-1279.

Vinod, P. and Sheeja, S. (2023). Sentiment prediction model in social media data using beluga dodger optimization-based ensemble classifier. *Social Network Analysis and Mining*, 13(1):1-16.

Wang, B. and Lu, W. (2018). Learning latent opinions for aspect-level sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2018)*, pp. 5537-5544, 2018.

Wang, X., Zhou, T., Wang, X., et al. (2022). Harshness-aware sentiment mining framework for product review. *Expert Systems with Applications*, 187:1-14.

Wang, Z., Zhu, X., Zhang, Y., et al. (2020). Sentiment forecasting in dialog. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages2448-2458.

Wen, J., Jiang, D., Tu, G., Liu, C., Cambria, E. (2023). Dynamic interactive multiview memory network for emotion recognition in conversation. *Information Fusion*, 91:123-133.

Wu, Z., Tang, S., Feng, X., Zou, J., Yin, F. (2023). Syntax-based aspect sentiment quad prediction by dual modules neural network for Chinese comments. *Computers, Materials and Continua*, 75(2): 2873-2888.

Zhang, S., Xu, H., Zhu, G., et al. (2022). A data processing method based on sequence labeling and syntactic analysis for extracting new sentiment words from product reviews. *Soft Computing*, 26(2): 853-866.

Zhen, Y., Luo, J.-D., Chen, H. (2023). Prediction of academic performance of students in online live classroom interactions—An analysis using natural language processing and deep learning methods. *Journal of Social Computing*, 4(1):12-29.

Zhou, H., Zheng, C., Huang, K., et al. (2020). KdConv: a Chinese multi-domain dialogue dataset towards multi-turn knowledge-driven conversation. In *Proceedings of 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, pages 7098-7108.

Zhou, Z., Tang, W., Li, M., Cao, W., Yuan, Z. (2023). A novel hybrid intelligent SOPDEL model with comprehensive data preprocessing for long-time-series climate prediction. *Remote Sensing*, 15(7):1-29.

Zou, C., Yin, Y., Huang, F. (2022). Attention-based emotion-assisted sentiment forecasting in dialogue. In *Proceedings of 2022 International Joint Conference on Neural Networks* (*IJCNN 2022*), IEEE, pages1-8.