Towards Strategic Persuasion: Unveiling Users' Susceptibility to Persuasive Strategies in Dialogues

Anonymous ACL submission

Abstract

Generative AI's rapid evolution has made dialogue systems indispensable tools. While persuasive strategies have been incorporated in dialogue systems to provide personalized services, current research primarily focuses on studying persuasive strategies from persuader's perspective, with limited exploration of persuadee's susceptibility towards these strategies. To bridge this gap, we introduce a novel task called Susceptibility Strategy Detection, aimed at identifying the persuasive strategies that users are most susceptible to. To support this new task, we develop a refined dataset P4G+, and propose a dual attitude-sensitive framework to detect susceptibility strategy by analyzing the persuasive process, user interactions, and content within dialogues. Comprehensive experiments have demonstrated the efficacy of our approach in identifying users' susceptible strategies. The code and dataset will be made available upon acceptance of this paper.

1 Introduction

In recent years, significant advancements in generative artificial intelligence (AI) have led to the emergence of intelligent dialogue systems. These systems offer personalized services by understanding user needs through conversational interactions. To get a more human-like service, efforts have been made to explore the persuasive ability of dialogue systems. This involves endowing AI with the capacity to understand users' preferences and adapt persuasive techniques accordingly. However, users' sensitivity (i.e., acceptance) to different persuasive strategies varies with factors such as age, gender, and personality (Chen et al., 2021). Mensah et al. referred to this sensitivity as "susceptibility," and the persuasive strategy that users are more likely to accept is termed their " susceptibility strategy" (Mensah et al., 2019). As shown in Figure 1, the persuader uses different persuasive strategies, and the Evidence strategy is not effective, whereas the Emotion strategy resonates more with the persuadee. Studying user susceptibility strategies can facilitate dialogue systems to provide more accurate and user-friendly persuasive interactions. However, current research primarily focuses on persuasion prediction (Wiegmann et al., 2022; Khatib et al., 2020), argument mining (Shmueli-Scheuer et al., 2019), and persuasive strategy identification (Iyer and Sycara, 2019; Kumar et al., 2023), while users' susceptibility to persuasive strategies has not thoroughly been investigated.

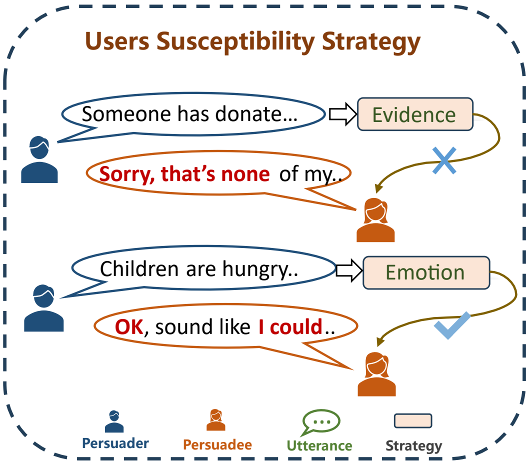

Figure 1: Description of User Susceptibility Strategy

To this end, this paper introduces a novel task called Susceptibility Strategy Detection, which can be regarded as a multi-turn dialogue understanding task. It aims to identify the user's susceptible strategies by analyzing the persuasive process, persuasive content, and users' interaction within the dialogue. Existing multi-turn dialogue modeling methods can mainly be divided into sequence-based and graph-based methods. Sequence-based methods treat the dialogue as a sequence and employ sequential models such as Long Short-Term Memory networks (LSTMs) to capture temporal and contextual information. However, these methods often struggle with long-range dependencies and have difficulties in capturing complex relationships across different elements of the dialogue. In contrast, graph-based methods exhibit better flexibility and stronger expressive power. These methods consider each utterance as a graph node and establish edges to depict the underlying dependency relationships between them, and employ techniques like graph neural networks (GNNs) to learn the structured information from the graph. Nonetheless, existing graph-based approaches haven't considered the distinct cognitive and behavioral patterns between persuaders and persuadees. Besides, the key to identifying user's susceptible strategies lies in accurately capturing user's attitude signals. Existing graph-based methods generally extract contextual semantic information, but fail to capture the subtle attitude-related information, such as emotional tendency and stance declaration. To address these limitations, first, we introduce a strategy-aware dialogue graph and a similar user graph, aimed at enhancing users representation and capturing the nuanced differences between them. Second, we propose an emotion-stance enhanced semantics extraction method to capture the subtle attitude shifts. In conclusion, our contributions are as follows:

- **Introduction of a New Task**: We introduce a new task of "Susceptibility Strategy Detection" in the context of persuasive dialogues. This task aims to identify the persuasive strategies that users are susceptible to, opening up new research opportunities in this field.
- **Development of a Dataset**: To support the new task, we develop a refined dataset P4G+, built upon the PersuasionforGood[1] through reannotation of the persuasive strategy and manual annotation of susceptibility labels.
- **Craft a Strategy-Aware Dialogue Graph**: We design a dialogue graph that incorporates strategy dependencies, customized for the susceptible strategy task, enabling a deeper understanding of the strategy's impact on the persuaded.
- **Proposal of Attitude-sensitive Framework**: We propose a Dual Attitude-sensitive framework for Susceptibility Strategy detection (DASS). The framework captures users' attitude shifts from emotion and stance aspects, leveraging the strategy-aware dialogue graph and a dual-channel GCN to jointly model dialogue processes and content. Comprehensive experiments demonstrate the effectiveness of our approach in identifying susceptibility strategies.

[1]https://gitlab.com/ucdavisnlp/persuasionforgood

2 Related Work

Multi-turn dialogue modeling methods can be primarily categorized into sequence-based and graph-based modeling methods.

2.1 Sequence-based methods

Sequence-based dialogue modeling method arranges dialogue sentences sequentially and utilizes sequential models to simulate the dialogue process. Early research focused solely on the text. For instance, Lee and Dernoncourt employed CNN and RNN to model sequential text to capture the local and global features of the dialogue sequence separately. To alleviate the issue of long-range dependencies in RNNs models, Dutt et al. utilized the GRU to capture temporal dependencies in dialogues and employed an attention mechanism to capture long-range dependencies. Previous research primarily concentrated on dialogue text and overlooked the unique language patterns of individual speakers. To address this limitation, certain researchers segmented the dialogue sequences according to the speakers and captured the distinct dialogue patterns of different roles (Hazarika et al., 2018b,a; Majumder et al., 2019). For instance, DialogueRNN (Majumder et al., 2019) employed independent GRUs to separately monitor the user's dialogue sequences and the global dialogue sequences, and subsequently integrated information between them through attention mechanisms.

However, when dealing with multi-turn persuasive dialogues, methods based on linear sequence structures struggle to express the various complex relationships within the dialogue. In contrast, graph structures offer a more robust and flexible expressive capacity through their ability to represent multiple elements and complex relationships using nodes and edges. Therefore, this paper employs graph-based approaches to tackle the challenge of susceptibility strategy detection.

2.2 Graph-based methods

Graph-based dialogue modeling methods transform dialogues into structured graphs and utilize graph neural network models to capture the information flow among nodes. Nodes and edges in the graph represent various elements and relationships of the dialogues. How to properly formalize dialogue scenes and represent them as the dialogue graph is the key challenge. Early work assumed influence exists between any two sentences and mapped the dialogue into a fully connected directed dialogue graph (Ghosal et al., 2019). The graph involves users' self-dependency and interactions, allowing nodes to access past and future dialogue content. Nevertheless, as future information is typically unavailable, Shen et al. constrained the edge directions to flow solely from the past to the future. They also introduced a window size to ensure that nodes only access the most recent historical nodes, thus alleviating redundancy. Lee and Choi extended the dialogue graph by introducing nodes for sentences, turns, subjects, and objects, along with edges for speakers, sentences, and arguments. They utilized GCN and multi-head attention mechanisms to capture contextual information of the dialogues. Previous methods focused on constructing nodes and edges based on dialogue utterances. To incorporate speaker information, Zhang et al. proposed depicting speakers as graph nodes and establishing a heterogeneous dialogue graph. However, diverse information from heterogeneous nodes might introduce ambiguity in semantics comprehension. To avoid information confusion in heterogeneous graphs, the DualGATs (Zhang et al., 2023) model constructed separate graphs for dialogue sentences and speakers, training them with graph attention networks (GATs) individually and merging the node features from both graphs with the attention mechanism.

Existing graph-based methods construct edges for different users using uniform rules, without considering the behavioral differences between persuaders and persuadees. Furthermore, most previous research initializes node embedding with utterance semantics extracted by text encoders like Bert(Devlin et al., 2019). However, identifying susceptibility strategy requires a detailed understanding of user's subtle attitude shifts. To address these limitations, a dual attitude-sensitive framework for susceptibility strategy detection is proposed.

3 Problem Definition

Given a dialogue dataset $D = \{d_1, d_2 \dots, d_n\}$ with n samples, each dialogue sample $d_i = \{(a_t^i, b_t^i)\}_{t=1}^{T}$ contains T pairs of sentences, where each pair (a_t^i, b_t^i) represents the sentences pair at turn t, with a_t^i representing persuader' sentence and b_t^i representing the persuadee's response. It is assumed that at least one persuasive strategy s_t is applied to each persuader's sentence a_t^i. The task of susceptibility strategy detection is to identify the persuasive strategies that the persuadee is susceptible to, which is formulated as a binary classification task, whose goal is to determine if the persuadee is persuaded at each turn t. If so, the strategies used by the persuader are considered as the susceptibility strategies for that particular persuadee.

4 P4G+ Dataset

To support our task, we construct P4G+ dataset based on PersuasionForGood (Wang et al., 2019). The data construction involves the annotation of persuasion strategies and susceptibility strategies.

4.1 Persuasion Strategy Annotation

Persuasion strategies are defined differently in various domains (Vargheese et al., 2020; Yang et al., 2019; Carlile et al., 2018). Chen and Yang summarize them and define eight more general persuasion strategies. To ensure data's applicability, we reannotate the persuasion strategies in P4G with these eight strategies. Detailed definitions of the strategies are presented in Appendix A.

We adopt the self-training paradigm (Nigam and Ghani, 2000) to annotate 10,170 instances of persuasion strategies, following the procedure below:
1). Manually annotate 2,100 instances of persuasion strategies to train a classifier.
2). Iteratively perform the following steps n times:
i. Predict 1,000 unannotated instances, verifying and correcting 60% of them.
ii. Merge all predictions into the training set to train a new classifier.
3). Utilize the latest classifier to predict the remaining unannotated samples.

In this paper, we employ LSTM as the classifier and iterate the process three times (n=3).

4.2 Susceptibility Strategy Annotation

We manually annotated 10,170 dialogue turns in the dataset for susceptibility strategy labels according to the following rule: For each dialogue turn, if there is a positive attitude shift of the persuadee, such as shifting from hesitation to affirmation, the susceptibility label for that turn is assigned as 1; otherwise, it is assigned as 0. In the dialogue turns labeled as 1, the persuasion strategy used by the current persuader is considered as the susceptibility strategy for the persuadee. Some annotated examples are provided in Appendix C.

A total of 1017 dialogue samples were annotated, excluding off-topic or meaningless dialogues, resulting in 807 valid samples. Within these valid samples, persuaders employed strategies a total of 9,205 times. According to Table 1, the Evidence strategy was the most frequently used, while the Reciprocity and Scarcity strategies were less common. The user's susceptibility strategies appeared 2,039 times in total, with the success rate of persuasive strategies generally ranging from 20% to 30%, which aligns with common intuition. Notably, the Politeness strategy showed a low success rate in persuasion. Possible reasons could be: first, the Politeness strategy is often used in initial greetings without intended persuasive effects; second, persuadees are typically not persuaded solely by polite language. Instead, a combination of Politeness and strategies such as Evidence would likely be more effective.

Following the data annotation and cleaning processes described above, P4G+ dataset is constructed, and its statistical information is presented in Table 2.

5 Methodology

In this section, we describe the proposed DASS model and detail the design of each module.

5.1 Dialogue Graph with Strategy dependency

In persuasive dialogue, the past persuasive effect of the strategies will influence the persuader's subsequent behavior. To model this influence, we introduce strategy dependency into the dialogue graph. The construction procedure of the dialogue graph $G = \{N_{g_i}, E_{g_i}\}$ for each sample is as follows:

Strategy	#Strategy	#Susceptible	Rate
Commitment	1279	292	22.8%
Emotion	1242	252	20.3%
Politeness	1366	64	4.7%
Reciprocity	298	81	27.2%
Scarcity	499	163	32.7%
Credibility	1260	346	27.5%
Evidence	2334	561	24.0%
Impact	927	280	30.2%
None	2396	273	11.4%

Table 1: Annotation Result of Susceptibility Strategy

Statistics	Value
# of dialogue samples	1017
# of valid samples	807
# of valid dialogue sentences	16140
# of total speakers	1285
Avg # of dialogue turns	10.01
Avg # of sentence words	17.17

Table 2: Statistics of P4G+

given a dialogue sample $d_i = \{u_1, u_2, \ldots, u_p\}$, each utterance u_i is taken as a node n_i, and the edges among nodes are construct based on self-dependency, inter-dependency, and strategy-dependency.

Self-dependency: In dialogue, a speaker's expression is influenced by their previous utterances. To depict this gradual development of personal discourse, we establish the self-dependency relationship by connecting the current node with the preceding W nodes of the same speaker.

Inter-influence: In persuasive scenarios, the persuader's statements directly impact the persuadee. Conversely, the persuadee's feedback reveals the extent of acceptance and attitude shifts towards the persuasive message. We model this interaction process by connecting the current sentence node n_i with the previous W sentence nodes uttered by the counterpart in the dialogue.

Strategy-dependency: We assume that if similar persuasion strategies are employed, there may exist some commonality or synergistic effect in terms of strategy. Therefore, we model strategy dependency by forming fully connected edges among persuader sentence nodes with the same persuasion strategy.

The window size W controls the maximum distance to prevent edge redundancy and reduce noise. In this paper, W is set to 3.

5.2 User Representation

Due to individual traits, different persuaders employing persuasive strategies have varying effects, so do the persuadees' sensitivity to these strategies. To characterize individual traits, this paper proposes an Attributes Enhancement via Group Augmentation method to enrich individual trait representation. The approach comprises two parts: feature extraction and feature enhancement with graph fusion.

We initialize users' features based on linguistic style and attribute information. Attribute in-

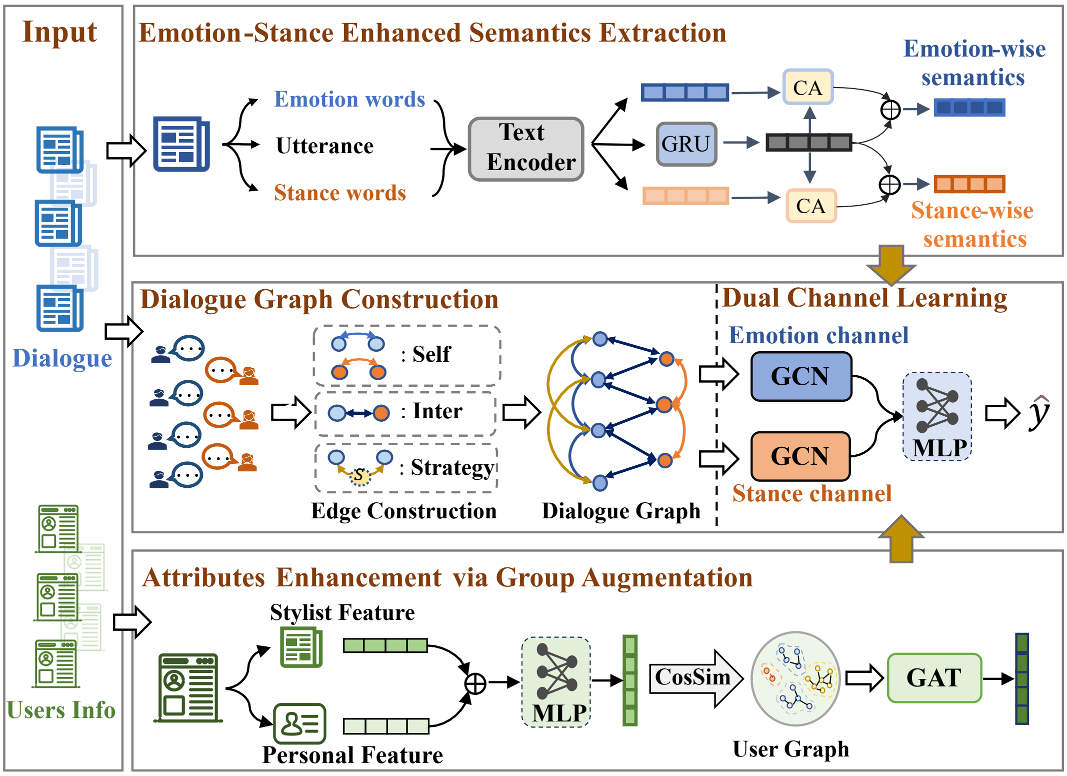

Figure 2: Framework of DASS

formation includes gender, education, personality scores, etc. From a linguistic perspective, we define 18 distinct styles from four aspects and extract these 18 style features from users' dialogue history. Detailed definitions are in Appendix B. To ensure consistency, we normalize the features to [0,1]. Consequently, we obtain attribute features h^i_{attr} and style features h^i_{sty}, which are then concatenated and projected into a high-dimensional space through a feedforward neural network, yielding the final initial user feature h_A, formalized as follow.

$$h_A = Tanh(W(h_{sty} \| h_{attr} + b) \tag{1}$$

We construct a similarity graph by connecting user nodes with similar traits. Each user is represented as a node in the graph, with an initial embedding h_A. Edges are established only when the cosine similarity between nodes embedding exceeds a certain threshold p. In this paper, p is set to 0.6.

With the user graph constructed, we apply Graph Attention Networks(Veličković et al., 2018) to aggregate the user representation with its associated group members adaptively. The final trait representation for user i is denoted as A_i

5.3 Emotion-Stance Enhanced Semantic Extraction

Capturing the users' attitude shifts within dialogue content is crucial for identifying susceptibility strategies. To achieve this goal, we introduce an emotion-stance enhanced semantic extraction method, involving word extraction and semantic enhancement.

Attitude-oriented Words Extraction

For a given dialogue sample $d_i = \{u_1, u_2, \dots, u_p\}$, where the i-th sentence is represented as $u_i = \{w_1, w_2, \dots, w_n\}$, we utilize the NLTK toolkit to perform part-of-speech tagging on sentence u_i to extract adjectives and adverbs as the emotion words $E_i = \{e_1, e_2, \dots, e_m\}$. Additionally, verbs and nouns are extracted as stance words $S_i = \{s_1, s_2, \dots, s_l\}$, where we assume verbs convey stance behaviors such as agreeing and nouns indicate the objects of stance like donations.

Emotion-Stance enhanced semantic extraction

Given utterance u_i, its emotion words E_i and stance words S_i, We employ a text encoder[2] to obtain the initial representations X_u, X_e, and X_s.

[2]In this paper, we initiate the word embedding by Glove 6B (https://github.com/stanfordnlp/GloVe)

First, we utilize a two-layer GRU to learn a sentence representation with context information, denoted as R_u. Then, we employ a cross-attention mechanism to extract the importance of emotion and stance words across the sentence. The cross-attention can be described as:

$$CA(Q, K, V) = softmax(Q' \cdot K' / \sqrt{d})V', \quad (2)$$

where $Q' = W_Q Q$, $K' = W_k K$, $V' = W_V V$ and d is the dimensionality. Given the representations X_u, X_e, and X_s, the cross-attention process is:

$$\begin{aligned} R_{e \to u} &= AvgPool(CA(X_e, R_u, R_u)), \\ R_{s \to u} &= AvgPool(CA(X_s, R_u, R_u)), \end{aligned} \quad (3)$$

where $AvgPool(\cdot)$ performs average pooling over the token representations generated by cross-attention to obtain one-vector text representations.

Finally, we obtain the emotion-stance enhanced semantic representation as follows:

$$\begin{aligned} R_e &= R_{e \to u} + R_u, \\ R_s &= R_{s \to u} + R_u, \end{aligned} \quad (4)$$

where the operator + denotes the element-wise addition of the vectors.

5.4 Dual Channel Learning & Classification

To better capture attitude signals, we propose a dual-channel graph convolutional network to jointly model dialogue processes and content. Given a dialogue sample d_i, we construct its corresponding dialogue graph $g_i = (N_{g_i}, E_{g_i})$. Node embeddings are initialized by concatenating individual trait representations with emotion-enhanced and stance-enhanced semantic representations.

$$\begin{aligned} h_i^e &= (R_e || A_u), \\ h_i^s &= (R_s || A_u). \end{aligned} \quad (5)$$

Subsequently, dual Graph Convolutional Networks (GCNs) are employed in parallel to capture emotion-specific and stance-specific information across separate channels. The update process of GCN is as follow:

$$\boldsymbol{H}^{(l+1)} = \sigma\left(\widetilde{\boldsymbol{D}}^{-\frac{1}{2}} \widetilde{\boldsymbol{A}} \widetilde{\boldsymbol{D}}^{-\frac{1}{2}} \boldsymbol{H}^{(l)} \boldsymbol{W}^{(l)}\right) \quad (6)$$

where $\widetilde{\boldsymbol{A}}$ is the adjacency matrix with self-loops, $\widetilde{\boldsymbol{D}}$ is the diagonal degree matrix, $\widetilde{\boldsymbol{D}}^{(l)}$ denotes the trainable weight matrix for the l layer, and $\sigma(\cdot)$ represents a nonlinear activation function.

After the update via GCN, let H_t^e, H_t^s respectively represent the final output of the emotion and stance channel for the persuadee at turn t. We employ different parametrized MLPs (multi-layer perceptions) to learn the attitude changes from emotion and stance perspectives between the previous and current turn. The procedure is as follows:

$$\begin{aligned} H_t^e &= \text{MLP}\left(H_{t-1}^e || H_t^e\right) \\ H_t^s &= \text{MLP}\left(H_{t-1}^s || H_t^s\right) \\ H_t^a &= H_t^e || H_t^s \end{aligned} \quad (7)$$

Finally, we apply an MLP as the predictor fed with H_t^a to predict the susceptibility label $\hat{y} = \{0, 1\}$ at turn t.

$$\hat{y} = sigmoid\left(\text{MLP}\left(H_t^a\right)\right) \quad (8)$$

$$L_{CE} = CrossEntropy(\hat{y}, y) \quad (9)$$

where $L_{CE} = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$ is a cross-entropy loss.

6 Experiments

In this section, we outline the baselines and evaluation metrics, and provide a thorough analysis of the experimental results and ablation study.

6.1 Comparison Methods

For a comprehensive evaluation, we compared our model with the following baselines, including sequence-based models and graph-based models. All the baselines are experimented and fine-tuned on the P4G+ dataset.

Sequence-based models: 1).LSTM (Hochreiter and Schmidhuber, 1997) is a classic variant of RNN that commonly serves as a benchmark model in sequential tasks. 2).DialogueRNN (Majumder et al., 2019) keeps track of individual party states with different GRUs, which enhances conversation understanding. 3). COSMIC (Ghosal et al., 2020) incorporates different elements of commonsense knowledge to effectively model interactions between speakers. 4).DialogueEIN (Liu et al., 2022) models intra-speaker, inter-speaker, global, and local emotional interactions, providing an understanding of emotional evolution in dialogues.

Graph-based models: 1).KET(Zhong et al., 2019) proposes a hierarchical self-attention to interpret utterance's context and incorporates external commonsense knowledge into utterances

Class	Model	Accuracy	Precision	Recall	F1	AUC
Sequence Based	LSTM	0.8175	0.5000	0.3082	0.3814	0.7954
	DialogueRNN	0.7025	0.3143	0.5517	0.4005	0.6437
	COSMIC	0.7944	0.4296	0.4310	0.4303	0.6526
Graph Based	KET	0.7298	0.3499	0.5828	0.4373	0.7493
	DialogueGCN	0.8225	0.5139	0.5068	0.5103	0.8273
	RGAT	0.8375	0.6034	0.4545	0.5185	0.8336
	DAG-ERC	0.8369	**0.6136**	0.4286	0.5047	0.8610
	DialogueEIN	0.8311	0.5326	0.5069	0.5194	0.8635
	DualGATs	0.8267	0.5178	0.5517	0.5342	0.8260
	DASS(Ours)	**0.8763**	0.6111	**0.6063**	**0.6087**	**0.8800**
	Improv.	3.88%	-0.25%	2.35%	7.45%	1.65%

Table 3: The table presents the results of DASS and other baselines for detecting Susceptibility Strategies on the P4G+ dataset. The best results are highlighted in bold, and improvements over the baselines are shown in red.

through a graph attention mechanism. 2).DialogueGCN(Ghosal et al., 2019) is a Graph Convolutional Neural Network that models conversational context by constructing a fully connected dialogue graph. 3). RGAT(Ishiwatari et al., 2020) considers self- and inter-speaker dependencies in conversations and enhances the graph-based neural network approach by incorporating relational position encodings. 4).DAG-ERC(Shen et al., 2021) proposes a Directed Acyclic Graph Network to model relationships among utterances and combines GNN models and RNN models to capture the temporal and context information of the dialogue . 5).DualGATs(Zhang et al., 2023) incorporates two individual GATs to analyze the complementary aspects of discourse structure and speaker information.

6.2 Evaluation Metrics

Susceptibility strategy detection can be regarded as a binary classification task for each dialogue turn. In this paper, five evaluation metrics are applied, including accuracy, precision, recall, F1 score, and AUC to assess the experimental results.

6.3 Results Analysis

The performance of our model and the compared methods are presented in Table 3. According to the results, it is observed that graphs-based methods generally outperform the sequence-based methods, demonstrating the advancement of the graph-based dialogue modeling approach. Our DASS model achieves the best performance in terms of Accuracy, Recall, F1 score, and AUC, with improvements of 3.88%, 2.35%, 7.45%, and 1.65% compared to the best results of the baselines, validating the effectiveness of our model. However, the DASS model is slightly lower in terms of the Precision metric compared to the DAG-ERC model, ranking the second-best performance. This may be attributed to the directed graph (DAG) modeling approach, which is beneficial for capturing temporal relationships in dialogues and contributes to precision. Similarly, the RGAT model with positional encoding also demonstrates good precision. In contrast, while the undirected dialogue graph used in this paper reduces the model's perception of temporal relationships, the introduced self, inter, and strategy relationships comprehensively enhance the model's ability to understand the dialogue context.

Among graph-based models, those integrating users' representations, namely DualGATs and our DASS model, have a better performance than models solely relying on dialogue semantic representations, namely DialogueGCN, RGAT, DAG-ERC, and DialogueEIN. This emphasizes the benefit of leveraging users' characteristics for detecting susceptibility strategy. Furthermore, our DASS model, in comparison with DualGATs, achieves better results by considering emotion and stance information, leading to a more precise capture of attitude shifts and yielding improved experimental results. Additionally, the RGAT model, which incorporates positional encoding in graph modeling, outperforms the DialogueGCN model, indicating the significance of addressing the lack of sequential information in graph-based methods.

6.4 Ablation Study

Ablation experiments were conducted to validate the effectiveness of the modules in our DASS model. According to the results shown in Table 4, three modules proposed in this paper were analyzed through ablation experiments.

Model	Accuracy	F1	AUC
DASS w/o Et	0.8375	0.4961	0.8448
DASS w/o St	0.8288	0.4830	0.8387
DASS w/o Et&St	0.8263	0.4755	0.8423
DASS w/o Grp	0.8438	0.5283	0.8612
DASS w/o Attr	0.8375	0.4758	0.8421
DASS w/o Stra	0.8375	0.5221	0.8695
DASS w/o Self	0.8375	0.5038	0.8627
DASS w/o Inter	0.8475	0.5987	0.8694
DASS r.p FCG	0.8438	0.5734	0.8666
DASS r.p DAG	0.8375	0.5221	0.8695
DASS	**0.8763**	**0.6087**	**0.8800**

Table 4: Ablation study on three modules of DASS. The first group is the Emotion-Stance enhanced module, the second group is the user's representation module, and the third group is the dialogue graph module.

Specifically, the Emotion-Stance enhanced module was assessed under three conditions: DASS w/o Et, DASS w/o St, and DASS w/o E&S, representing the model without emotion-enhanced embedding, stance-enhanced embedding, and both. Compared to DASS model, these variations resulted in performance reductions of 9.65%, 6.06%, and 9.85% respectively in terms of F1, demonstrating the effectiveness of the emotion-stance enhanced module. For the user's characteristic representation module, we developed DASS models without group embedding (DASS w/o Grp) and without all user embeddings (DASS w/o All), resulting in performance decreases of 10.49% and 12.48% in F1 score, respectively.

For the user's representation module, we developed DASS models without group embedding (DASS w/o Grp) and without all user embeddings (DASS w/o Attr), resulting in performance decreases of 10.49% and 12.48% in F1 score, respectively.

For the dialogue graph module, we conducted dependency-level and graph-level ablation experiments to assess the proposed dialogue graph. In the dependency-level ablation, we introduced DASS models without strategy dependency (DASS w/o Stra), self-dependency (DASS w/o Self), and inter-influence (DASS w/o Inter). These models all demonstrated varying degrees of performance decrease, highlighting the effectiveness of the three types of dependency relationships in this study. Notably, the DASS w/o stra model exhibited the most significant performance decrease, with respective drops of 8.83% and 1.74% in F1 score and AUC, highlighting the crucial role of strategic dependency in detecting susceptibility strategies. In the graph-level ablation, we replaced the dialogue graph of our DASS model with a fully-connected graph (FCG) used in DialogueGCN, and a directed acyclic graph (DAG) used in DAG-ERC, namely DASS r.p FCG and DASS r.p DAG. The DASS r.p FCG model, while able to comprehensively describe various dependencies in the dialogue, introduced redundant information due to excessive edge relationships, resulting in lower performance compared to the DASS model. Moreover, the DASS r.p DAG model contained self-dependencies of the speakers and temporal dependencies of the sentences, which was consistent with the DASS w/o strategy model and had similar experimental results, further validating the effectiveness of the strategy dependency in this paper.

7 Conclusion

In this paper, our primary contribution lies in the proposal of a new task and the development of a solid method to address it. Specifically, we introduce the task of detecting susceptibility strategies, develop the corresponding P4G+ dataset through the reannotation of persuasive strategies and the manual annotation of susceptibility labels. Our method integrates a strategy-aware graph to analyze dialogue flow, an attitude-sensitive module for content semantics extraction, and speaker representations augmented with group attributes. Extensive experiments are conducted to assess our model's efficacy, in comparison to established sequence- and graph-based models. Results show that our model achieves competitive performance.

8 Limitations

While this paper introduces a new task dataset and a corresponding solution method, they both have some limitations.

- In terms of the practicality of the task, we assume that the historical dialogues of existing users all contain persuasive strategies that have been used, and we identify the susceptible strategies of users from them. Consequently, our task framework may not identify susceptible strategies for new users without historical dialogues or in cases where the strategies used in the dialogue are not explicitly stated.

- As for the dataset, we develop P4G+ dataset based on the PersuasionforGood dataset, which is a persuasion dialogue dataset in the donation domain. Therefore, there is a need to further expand the dataset in terms of quantity and domain, proposing a more comprehensive and multidimensional dataset for a more thorough evaluation.

- In terms of the method, the model in this paper is based on the graph method. While it offers stronger expressive power compared to sequential models, it naturally lags in capturing temporal information in dialogues. Consequently, in precision metric, one sequence-based model even outperforms the proposed method. Therefore, enhancing the graph-based method's capability to capture sequential temporal information is one of the key directions for future improvement.

References

Winston Carlile, Nishant Gurrapadi, Zixuan Ke, and Vincent Ng. 2018. Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 621–631. Association for Computational Linguistics.

Hui Chen, Deepanway Ghosal, Navonil Majumder, Amir Hussain, and Soujanya Poria. 2021. Persuasive dialogue understanding: The baselines and negative results. *Neurocomputing*, 431:47–56.

Jiaao Chen and Diyi Yang. 2021. Weakly-supervised hierarchical models for predicting persuasive strategies in good-faith textual requests. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 12648–12656. AAAI Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *Preprint*, arXiv:1810.04805.

Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Haogang Bao, and Carolyn P. Rosé. 2021. Resper: Computationally modelling resisting strategies in persuasive conversations. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 78–90. Association for Computational Linguistics.

Deepanway Ghosal, Navonil Majumder, Alexander F. Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. COSMIC: commonsense knowledge for emotion identification in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*, volume EMNLP 2020 of *Findings of ACL*, pages 2470–2481. Association for Computational Linguistics.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander F. Gelbukh. 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, Hong Kong, China, November 3-7, 2019*, pages 154–164. Association for Computational Linguistics.

Devamanyu Hazarika, Soujanya Poria, Rada Mihalcea, Erik Cambria, and Roger Zimmermann. 2018a. ICON: Interactive conversational memory network for multimodal emotion detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2594–2604, Brussels, Belgium. Association for Computational Linguistics.

Devamanyu Hazarika, Soujanya Poria, Amir Zadeh, Erik Cambria, Louis-Philippe Morency, and Roger Zimmermann. 2018b. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 2122–2132. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Taichi Ishiwatari, Yuki Yasuda, Taro Miyazaki, and Jun Goto. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 7360–7370. Association for Computational Linguistics.

Rahul Radhakrishnan Iyer and Katia P. Sycara. 2019. An unsupervised domain-independent framework for automated detection of persuasion tactics in text. *CoRR*, abs/1912.06745.

Khalid Al Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. 2020. Exploiting personal characteristics of debaters for predicting persuasiveness. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7067–7072. Association for Computational Linguistics.

Yaman Kumar, Rajat Jha, Arunim Gupta, Milan Aggarwal, Aditya Garg, Tushar Malyan, Ayush Bhardwaj, Rajiv Ratn Shah, Balaji Krishnamurthy, and Changyou Chen. 2023. Persuasion strategies in advertisements. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 57–66. AAAI Press.

Bongseok Lee and Yong Suk Choi. 2021. Graph based network with contextualized representations of turns in dialogue. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 443–455. Association for Computational Linguistics.

Ji Young Lee and Franck Dernoncourt. 2016. Sequential short-text classification with recurrent and convolutional neural networks. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 515–520. The Association for Computational Linguistics.

Yuchen Liu, Jinming Zhao, Jingwen Hu, Ruichen Li, and Qin Jin. 2022. Dialogueein: Emotion interaction network for dialogue affective analysis. In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 684–693. International Committee on Computational Linguistics.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander F. Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive RNN for emotion detection in conversations. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6818–6825. AAAI Press.

Humphrey Mensah, Lu Xiao, and Sucheta Soundarajan. 2019. Characterizing susceptible users on reddit's changemyview. In *Proceedings of the 10th International Conference on Social Media and Society*, pages 102–107.

Kamal Nigam and Rayid Ghani. 2000. Analyzing the effectiveness and applicability of co-training. In *Proceedings of the 2000 ACM CIKM International Conference on Information and Knowledge Management, McLean, VA, USA, November 6-11, 2000*, pages 86–93. ACM.

Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. 2021. Directed acyclic graph network for conversational emotion recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 1551–1560. Association for Computational Linguistics.

Michal Shmueli-Scheuer, Jonathan Herzig, David Konopnicki, and Tommy Sandbank. 2019. Detecting persuasive arguments based on author-reader personality traits and their interaction. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2019, Larnaca, Cyprus, June 9-12, 2019*, pages 211–215. ACM.

John Paul Vargheese, Matthew Collinson, and Judith Masthoff. 2020. Exploring susceptibility measures to persuasion. In *Persuasive Technology. Designing for Future Change - 15th International Conference on Persuasive Technology, PERSUASIVE 2020, Aalborg, Denmark, April 20-23, 2020, Proceedings*, volume 12064 of *Lecture Notes in Computer Science*, pages 16–29. Springer.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. *Preprint*, arXiv:1710.10903.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 5635–5649. Association for Computational Linguistics.

Matti Wiegmann, Khalid Al Khatib, Vishal Khanna, and Benno Stein. 2022. Analyzing persuasion strategies of debaters on social media. In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 6897–6905. International Committee on Computational Linguistics.

Diyi Yang, Jiaao Chen, Zichao Yang, Dan Jurafsky, and Eduard H. Hovy. 2019. Let's make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms.

In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 3620–3630. Association for Computational Linguistics.

Dong Zhang, Liangqing Wu, Changlong Sun, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5415–5421. ijcai.org.

Duzhen Zhang, Feilong Chen, and Xiuyi Chen. 2023. Dualgats: Dual graph attention networks for emotion recognition in conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 7395–7408. Association for Computational Linguistics.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. 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, Hong Kong, China, November 3-7, 2019*, pages 165–176. Association for Computational Linguistics.

Appendix

A Strategy Definition

Table 5 presents the detailed definitions and corresponding examples of the eight general persuasion strategies defined in (Chen and Yang, 2021)

Strategy	Definition & Instance
Commitment	The persuaders indicating their intentions to take acts or justify their earlier decisions to convince others that they have made the correct choice e.g., *I have donated money to this institution, and it turns out I was right.*
Emotion	Making request full of emotional valence and arousal affect to influence others. e.g.,*I've been in the lowest depressive state of my life.*
Politeness	The usage of polite language in requests. e.g.,*Your help is deeply appreciated!*
Reciprocity	Responding to a positive action with another positive action. People are more likely to help if they have received help themselves. e.g.,*I'll pay it forward with my first check.*
Scarcity	People emphasizing on the urgency, rare of their needs. e.g.,*I haven't eaten a meal in two days.*
Credibility	The uses of credentials impacts to establish credibility and earn others' trust. e.g.,*I can provide any documentation needed.*
Evidence	Providing concrete facts or evidence for the narrative or request. e.g.,*My insurance was canceled today.*
Impact	Emphasizing the importance or impact of the request. e.g.,*This loan will help him with his business.*

Table 5: Eight general persuasion strategies defined in the reference (Chen and Yang, 2021)

B The Proposed 18 Stylistic Feature

Table 6 presents the 18 stylistic features proposed in this paper. We extract the user's stylistic feature from four aspects, including Lexical Usage, Language Preference, Perspective Focus, and Tonal Style. We use the ratio of each aspect as the field. For example, the verb ratio represents the proportion of the number of verbs to the total number of words.

Feature Category	Specific Fields
Lexical Usage	Preposition ratio, Verb ratio, Noun ratio, Adjective ratio Adverb ratio, Average word length
Language Preference	Absolute word ratio, Ambiguous word ratio Number ratio, Uppercase word ratio
Perspective Focus	First-person pronoun ratio, Second-person pronoun ratio Third-person pronoun ratio, Pronoun ratio
Tonal Style	Punctuation ratio, Pause ratio Exclamation mark ratio, Sentence period ratio

Table 6: Linguistic style feature

C Annotated Examples of P4G+

Here present two annotated examples of the P4G+ dataset. Each example consists of a dialogue sample spanning 10 turns. In each turn, the first sentence is spoken by the persuader, and the second sentence is spoken by the persuadee. The strategy employed by the persuader is highlighted in red. The label indicates whether the persuader has successfully persuaded or changed the attitude of the persuadee. For instance, in the first sample, the persuader successfully changed the attitude of the persuadee from "not ready to donate" in the 6^{th} turn to "OK, I will donate" in the 7^{th} turn with the strategy of Impart. Therefore, the Impart strategy is regarded as the Susceptibility Strategy of the persuadee.

Turn	Conversation	label
1	Hello (None) Hello, how are you?	0
2	I am well. I was wondering......do you often donate to charity? (None) That's great! I actually do not donate to charity often, but I believe I will in the future.	0
3	Have you ever heard of Save the Children Foundation? (None) Yes, I have heard of the Save the Children Foundation.	0
4	Would you be interested in donated part of your task payment directly to Save the Children? (None) I actually am not ready to donate yet. I would like to find out more before donating.	0
5	Millions of Syrian children were killed or injured. You can learn more at URL (Evidence, Impact) Thank you for providing additional information, but I am not ready to donate yet.	0
6	Just .77 a day can help shelter a refugee family. That's less than a cup of coffee (Impact) I understand, but I need more time to think about it. I am not ready to donate right now.	0
7	You could make just a one-time donation today. Only .10 would change a life. (Impact) Okay, I will donate .10 today.	1
8	Thank you! Your donation is greatly appreciated and will change a child's life! (Emotion, Politeness) You are welcome. I hope that I will be able to donate more in the future.	0
9	I do as well. Have a wonderful day! (Emotion, Politeness) You too! Have a wonderful day!	0
10	Thank you very much!(Politeness) Thank you for taking time to share so much information about the Save The Children Foundation.	0
Susceptibility Strategy: Impact		

Turn	Conversation	label
1	Hi, would you like to donate to Save the Children? (None) Hello, I have heard of the organization but I don't know what would my donation go toward?	0
2	Your donation would help children in developing countries. It will make a difference.(Evidence, Impact) I would rather help to an American children.	0
3	Are American children more important than non-American children? (None) I am American and feel it important to help my fellow Americans	0
4	The children we help are in desperate need. Can't you spare even a small donation?? (Scarcity) Perhaps I could help out a little. Where are you at?	1
5	You can check us out at URL(Evidence) are you based in America? are you an American?	0
6	We're an international NGO based in the UK.(Evidence) looking at the link, it appears that children in America also benefit from your charity. Is that correct?	0
7	Ah, so you noticed. (Impact) I hate to be so prejudicial, but I care about the kids around me more than ones I cannot see!	0
8	Well, then. You are a scallywag ninny! (Emotion, Politeness) I accept that! How much of a donation are you looking for from me?	0
9	Your donation will be deducted from your task payment. Choose any amount you want. (None) all of it seems a little steep! I would consider a smaller portion. half would work for me.	1
10	Well, it's something. Let's do it, then. Half. Deal? (None) yes, fifteen cents	0
Susceptibility Strategy: Impact		