

RESEARCH ARTICLE

A personalized model for joint dialogue emotion classification and act recognition

Zhiqiang Ma^{1,3*} , Yutong Zhou² , Biqi Xu¹ , Kai Lv¹ , Jia Liu¹ 

¹Department of Intelligent Science and Technology, College of Data Science and Application, Inner Mongolia University of Technology, Hohhot, Inner Mongolia, China

²Department of Information Engineering, Xuchang Electric Vocational College, Xuchang, Henan, China

³Department of Intelligent Science and Technology, Inner Mongolia Autonomous Region Engineering & Technology Research Center of Big Data Based Software Service, Inner Mongolia University of Technology, Hohhot, Inner Mongolia, China

Abstract

Background: Joint dialogue emotion classification and act recognition aim to simultaneously predict emotional and behavioral labels for each utterance. While existing models leverage discourse context, they largely overlook individual differences in sequential expression patterns.

Objective: This paper proposes a personalized model integrating personality for joint dialogue emotion classification and act recognition (PEDR-GAT) to incorporate personality traits into both local semantics and global temporal reasoning, thereby bridging inference gaps caused by neglected individual differences.

Methods: First, personality information and label distributions were used to initialize utterance representations, strengthening the association between personality-guided semantics and task labels. Second, graph attention mechanisms extracted global temporal dependencies across utterances and captured cross-task interactions. Finally, a co-attention mechanism integrated personality-guided local and global features to maintain long-term dependencies and supplement short-term contextual information.

Results: Experiments on Mastodon, DailyDialog, and CPED datasets show that PEDR-GAT outperforms baselines, achieving F1 improvements of 2.0%/0.6%, 2.3%/1.9%, and 5.3%/2.3% for emotion/act recognition, respectively. Personality-guided semantic label acquisition enhances semantics-label associations, and personality-fused graph attention effectively captures latent temporal patterns, demonstrating the value of incorporating individual differences into joint dialogue modeling.

Conclusion: PEDR-GAT effectively integrates personality traits into joint dialogue emotion and act recognition, demonstrating that modeling individual differences significantly enhances performance across multiple datasets.

Keywords: Personality; Graph attention mechanisms; Dual-task recognition; Multi-head attention

*Corresponding author:

Zhiqiang Ma
(mzq_bim@imut.edu.cn)

Citation: Ma Z, Zhou Y, Xu B, Lv K, Liu J. A personalized model for joint dialogue emotion classification and act recognition. *Health Psychol Res.* 2026;14(2):0401.
doi: 10.36922/hpr.0401

Received: November 24, 2025

Revised: February 16, 2026

Accepted: February 27, 2026

Published online: May 25, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons AttributionNon-Commercial 4.0 International (CC BY-NC 4.0), which permits all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The joint recognition of dialogue emotion and act tasks involves the collaborative modeling of two tasks within dialogue: dialogue act recognition (DAR) and emotion recognition in conversation (ERC).^{1,2} These tasks share a potential emotional correlation, mutually

facilitating and executing each other. The underlying emotional link between these tasks is two-fold. On the one hand, DAR provides cues for emotion recognition (ERC), and, reciprocally, emotion transitions can improve the prediction of dialogue acts.^{3–5} In a conversation from the Chinese psychological emotion dialogue (CPED)⁶ dataset, as depicted in Table 1. When predicting person B's act, person B's agreement-act label suggests that both parties tend to share the same emotional label when expressing similar opinions. Furthermore, understanding emotional information helps predict current actions. The joint recognition of dialogue acts and emotions provides a more comprehensive grasp of the speaker's intention,⁷ contributing to machine recognition and understanding of the speaker's psychological state.⁸

Emotion perception aims to recognize emotion-related sequences through contextual analysis. Low recognition accuracy often leads to ineffective transmission of emotional information in related tasks such as emotion state generation and expression. Several disciplines, including cognitive psychology, emotion psychology, and neuropsychology, have demonstrated a close relationship between emotional state and cognition. Currently, cognitive reliance on behavior is common in autonomous emotion generation, and this also occurs in dialogue systems as conversational behavior. Existing research on joint emotion–behavior perception typically employs mechanisms such as context control selection and dual-network differentiation to perform emotion category recognition and behavior category recognition in dialogue. However, it overlooks the correlation between personality-related cues and the two associated sequences. Due to different upbringing environments, individuals exhibit individual differences in their emotional state sequences when facing the same environment. Personality traits are thus correlated with individual differences in emotion generation processes. In emotion perception, without personality-related cues, the perceived emotional sequences may show overly low discrete emotional correlations. Ultimately, the perceived emotional sequences show no individual differences within the same context.

As illustrated in Figure 1, different information concepts are abstracted in the context as nodes based on social psychology. Speaker personality correlates with

emotions and acts in specific contexts, with emotion and act labels providing useful references within the same personality. Different speakers in interactive processes exhibit explicit interaction dependencies in dialogue acts, and there exists a latent correlation between emotions. Leveraging dual-task dependencies enables the model to understand internal emotional and act variations within speakers with the same personality, facilitating a better focus on the potential correlations embedded in discourse.

The paper proposes a personalized model integrating personality for joint emotion classification and act recognition in dialogue. The aim is to enhance the accuracy of emotion recognition in conversation and dialogue act recognition by concurrently considering speaker-personality-aware discourse structures and the temporal interactions between the two tasks. This model comprises three hierarchical levels: a feature extraction initialization layer, a personality-fused interactive reasoning layer, and a dual-task classification layer. The feature extraction initialization layer leverages personality-guided discourse to perform initial local feature extraction across dual tasks. The personality-fused interactive reasoning layer is designed to capture semantic temporal correlations between dialogues under different personalities, as well as temporal associations both within and between the two tasks under different personalities. The dual-task classification layer integrates global discourse information and dual-task information under different personalities for classification. Figure 2 provides a simplified overview of the personalized model integrating personality for joint dialogue emotion classification and act recognition (PEDR-GAT). The model components are discussed in detail in Section 3.

The main contributions of this paper are as follows:

- We propose a personality-based model for joint emotion classification and act recognition. We introduce the use of the Big Five personality traits for the first time in joint emotion classification and behavior recognition.
- We propose a graph-based model that integrates personality-aware utterance temporal graphs and dual-task interaction temporal graphs, enhancing temporal dependencies in speech and interaction

Table 1. Emotion distribution in Chinese psychological emotion dialogue

Speaker	Utterances (U)	Dialogue act recognition label	Emotion recognition in conversation label
Person A	U1: I have only read three of them all online.	Statement	Happy
Person A	U2: This reading is also a super favorite of mine.	Agreement	Happy

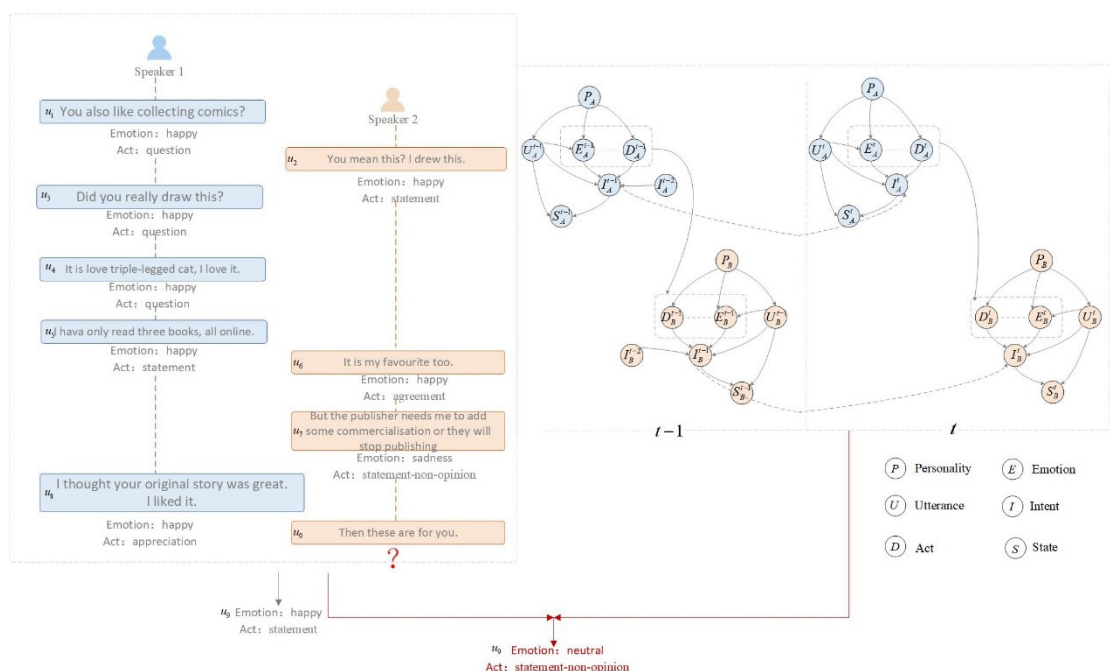


Figure 1. Emotion classification and act recognition are performed under the interaction of dialogue history and different variable factors, where P represents the speaker's Big Five personality traits, U represents the utterance, D represents the dialogue act, E represents the dialogue emotion, I represents the speaker's intent, and S represents the speaker's state. This provides a more precise understanding of the relationships between different tasks.

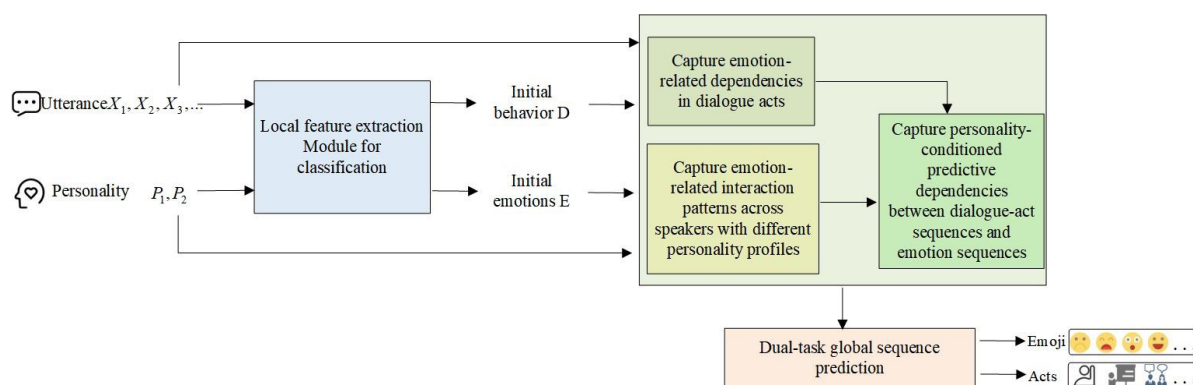


Figure 2. A simplified overview of the personalized model integrating personality for joint dialogue emotion classification and act recognition model architecture

dependencies in tasks among different speakers with the same personality.

- Experimental results indicate that PEDR-GAT achieves better emotion classification and personality-based behavior recognition in dialogues.

2. Related work

In the initial stages of various text recognition tasks, most research treated recognition as a standalone text classification task, using traditional statistical classifiers. With the advancement of recurrent neural networks,

contemporary recognition methods primarily fall into two categories: sequence labeling and graph neural networks.

2.1. Sequence labeling method

The sequence labeling method involves treating recognition as a sequence labeling task, addressing recognition challenges by training and memorizing encoded features within the sequence. Relying on sequence models enables more effective capture of temporal dynamics across sequences.

In the realm of sequence-labeling-based emotion recognition, Ayetiran *et al.*⁹ employ a combination of convolutional and bidirectional long short-term memory networks to jointly learn emotional data at both the document and aspect levels. Majumder *et al.*¹⁰ introduce the dialogue recurrent neural network (DialogueRNN) emotion recognition model, treating users and dialogue systems as independent entities. They dynamically capture local-level changes in user emotion and then use static global context modeling for emotion recognition. Inspired by emotion cognition theory, Hu *et al.*¹¹ propose the DialogueCRN model, integrating long short-term memory (LSTM) and attention mechanisms to identify emotional nuances in dialogues.

In the realm of sequence-labeling-based act recognition tasks, Kumar *et al.*¹² employ bidirectional time series and a random field to identify acts within dialogues. Raheja *et al.*¹³ use attention mechanisms and time-series (LSTM) models for action recognition. During the encoding phase, Duran *et al.*¹⁴ integrate act and sentence encoding using bidirectional encoder representations from transformers, systematically comparing and validating the performance differences among various encoding mechanisms in act classification. Malhotra *et al.*¹⁵ leverage gated recurrent units and a time-sliding window to perceive local and global contextual cues in dialogues for act recognition.

In the domain of multitask sequence labeling for emotion and act recognition, Cerisara *et al.*¹ introduce a multitask framework based on a shared encoder that implicitly models inter-task correlations. Kim *et al.*² devise a convolutional neural network that integrates dialogue act, predictive factors, and emotion recognition into a unified model. Qin *et al.*³ present a cooperative interaction layer to model the interplay between the two related tasks. Zheng *et al.*¹⁶ introduce a bidirectional multi-hop reasoning network that focuses on explicit correlations between acts and emotions to capture richer act and emotional cues, thereby enhancing effectiveness and accuracy in multitask recognition inference. Li *et al.*¹⁷ propose a dual-channel perception dynamic convolutional network to dynamically generate kernels and capture crucial local context.

2.2. Graph neural networks

Graph neural networks have garnered increasing attention in recognition tasks, leveraging their ability to refine dependency features of speakers or historical contexts more effectively.

In the realm of emotion recognition using graph neural networks, Ghosal *et al.*¹⁸ introduce the dialogue graph convolutional network (DialogueGCN) model. This model addresses the challenge of emotion differentiation during

context propagation by constructing a heterogeneous graph with speech and speakers as nodes and distinct connecting edges to differentiate speakers. Zhang *et al.*¹⁹ use a graph structure to represent sentences and speakers in the dialogue history as nodes, forming a dialogue graph by connecting them with edges that represent statements. Subsequently, they employ graph convolutional networks to extract information for emotion recognition. Zhao *et al.*²⁰ propose two distinct methods for executing context and speaker-specific modeling in separate threads. Zhang *et al.*²¹ structured the discourse and interaction between discourses into two graph structures, integrating information through a cross-attention mechanism for emotion recognition.

In the domain of joint recognition tasks involving both act and emotion using graph neural networks, Qin *et al.*²² apply graph attention networks to an undirected, disconnected graph composed of isolated, speaker-specific fully connected subgraphs, concurrently executing both tasks. Xing *et al.*⁵ propose the discourse-aware relational emotion reasoner (dialogue act and emotion recognition [DARER]) model, a graph-based structure that transforms contextual discourse relations and dual-task relations into a concatenated graph to learn context-, speaker-, and time-sensitive discourse representations. Given the confidence in embedded contextual discourse, Xu *et al.*⁷ propose a bidirectional learning-integrated model based on joint DAR and ERC. Li *et al.*²³ introduce a multi-information-aware recurrent reasoning network and design a speaker-time-aware heterogeneous graph for semantic-level interaction and self-interaction, thereby facilitating joint task recognition.

However, the above-mentioned methods overlook the correlation between the speaker's personality and both dialogue behavior and emotion. Simultaneously, it ignores the dependencies within and between tasks, as well as their relation to the speaker. This dependency manifests in the continuity of the same task across the same speaker and the interaction dependencies between different speakers in different dialogues. There is often an explicit correlation between categories across different tasks within the same speaker, and between different tasks across different speakers, with potential mapping relationships.

3. Research methodology

This section describes the classification model in detail. The frame of the model is shown in [Figure 3](#).

Firstly, the initiation of local discourse features and task-label feature representations was achieved through personality-guided label text acquisition. Secondly, the

directed temporal associations between personality information and discourse information, as well as between personality information and dual-task information, were established using graph neural networks, with a fusion mechanism that amalgamates the two. Lastly, the classifier produced recognition results based on the fused features.

Personality traits are correlated with how emotions are elicited and evolve over time. Our objective is to jointly infer emotion and dialogue act by modeling the complex interactions among utterances, speaker personality traits, emotions, and dialogue acts in conversation. To this end, we constructed three graph structures that captured dependencies from distinct perspectives. Across these three graphs, we incorporated personality traits in both the local feature initialization stage and the temporal graph reasoning stage for complementary reasons.

- Local feature initialization. Injecting personality traits at this stage aims to initialize each utterance with a more speaker-specific representation that reflects individual characteristics. We treated personality traits as an utterance-level anchor (rather than merely using a speaker identifier), providing a better starting point for subsequent deep modeling.
- Temporal graph reasoning. Injecting personality traits at this stage allows personality information to directly influence message passing and aggregation when modeling dynamic interactions and dependencies among utterances. Concretely, personality traits were

incorporated into graph nodes and/or edge-related factors, enabling them to modulate how information propagates over time and across interlocutors.

Overall, personality participation in these two stages was complementary: the initialization stage mainly benefits semantic understanding and personalization of utterance representations, while the temporal graph reasoning stage focused on personality-aware dynamic exchanges between emotion and dialogue act along the dialogue flow. The two components reinforced each other rather than duplicating the same effect. The following sections will introduce the model's structure in turn.

3.1. Problem analysis

Existing joint emotion and behavior recognition models overlook individual differences across speakers, resulting in a lack of uniqueness in dialogue emotion and behavior recognition and low recognition rates. The problem is formulated as follows. Given a dialogue $D = \{U_1, R_1, \dots, R_{T-1}, U_T\}$ consisting of N utterances, along with a set of accompanying speaker personalities $P = (p^U, p^R, p^U, \dots, p^U)$. Each utterance $U_i = (w_{i1}, w_{i2}, \dots, w_{iM})$ was a sequence of M words, with speaker personality $p^* = (\text{Neuroticism}, \text{Extraversion}, \text{Openness}, \text{Agreeableness}, \text{Conscientiousness})$. The objective of this paper is to model the dialogue emotion sequence $EmoS = (e_1^U, e_1^R, \dots, e_{T-1}^R, e_T^U)$ and the dialogue behavior sequence $ActS = (d_1^U, d_1^R, \dots, d_{T-1}^R, d_T^U)$ under different personality characteristics. This can be described by Equation 1.

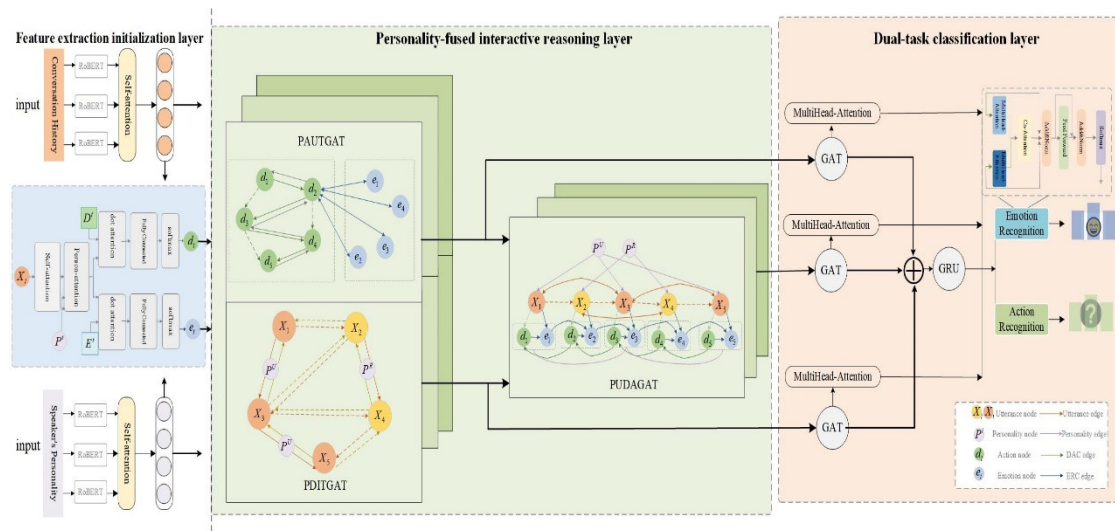


Figure 3. Personalized model integrating personality for joint dialogue emotion classification and act recognition (PEDR-GAT) model structure. Purple edges (directed: personality-utterance ownership; undirected: personality-dual-task ownership); orange edges (solid: intra-personality temporal flow; dashed: cross-personality utterance flow); blue edges (bidirectional: act-associated implicit emotions; unidirectional: utterance-emotion and emotion temporal relations); and green directed edges for act temporal flow (bidirectional to capture feedback dynamics).

Abbreviations: GAT: Graph attention network; GRU: Gated recurrent unit; PAUTGAT: Personality-aware utterance temporal graph; PDITGAT: Personality dual-task interaction temporal graph; PUDTAT: Personality-aware utterance and dual-task integrated temporal graph.

$$\text{IM}(C) \stackrel{\text{def}}{=} \arg \max_{\text{EmoS}, \text{ActS}} P(\text{EmoS}, \text{ActS} \mid D, P^U, P^R) \quad (1)$$

The focus of this task lies in: (i) modeling the association between personality traits dependent on speech and emotion and behavior; (ii) modeling the interaction between emotion and behavior. The challenge of this task lies in how to perceive the personality-dependent emotion and behavior sequences using limited contextual information in the dialogue.

3.2. Feature extraction initialization layer

This section provides a detailed exposition of the model that integrates personality information and task labels. The objective is to initialize the encoding of text sequences with labeled information, establishing a more intricate relationship between personality-guided labels and discourse (**Figure 3**). The model primarily comprises bidirectional encoder representations from transformer (BERT) semantics and label extraction, and multi-head attention layers.

For a given dialogue history $U = \{x_1, x_2, \dots, x_n\}$, where the i -th utterance was denoted as $x_i = (w_{i,1}, w_{i,2}, w_{i,3}, \dots, w_{i,M})$, the speaker's Big Five personality traits information, denoted as P^i , was appended. This results in the formation of $X = ([CLS_1], p_1, x_1, [CLS_2], p_2, x_2, [CLS_3], p_3, x_3, \dots, [CLS_N], p_N, x_N)$, and these elements were concatenated to serve as input to BERT, generating a contextual representation endowed with discourse and personality features.

This section focuses on extracting encoded information that integrates personality, emotions, and initial local features of the act. This process involved two branches: one encoded the initial local features of emotional aspects in dialogue, while the other encoded the initial local features of dialogue acts. The feature vector H of spoken discourse x_i undergoes a self-attention process, resulting in $H_x = \text{MAtt}(H, H, H)$, followed by guidance using personality features to obtain $H_{pe} = \text{MAtt}(H_x, H_p, H_p)$. Subsequently, H_{pe} was employed to calculate the initial distribution probabilities for both emotional and act aspects from the distributions of act and emotional discourse. The detailed step-by-step computations are provided in Section A1.1.

3.3. Personality-fused interactive reasoning layer

The personality-fused interactive reasoning layer employed three graph attention neural networks to effectively capture the temporal structure of speech perceived by the speaker's personality, the temporal structure of dialogue acts and emotional expressions, and the interactive structural relationships between the two tasks.

We acknowledged that the Big Five traits, emotions, and

dialogue acts might be conceptually correlated. Our graph-based model did not treat them as independent covariates in a linear regression that estimates separate linear effects. We represented personality as context nodes and used non-linear attention-based message passing to regulate relations and representations among utterance nodes. This design aimed to learn complex, non-linear interaction patterns among personality, emotion, and dialogue acts. While the non-linear attention mechanism mitigated the strict multicollinearity issues inherent in linear regression, we acknowledged that inter-trait correlations (e.g., the typical negative correlation between neuroticism and extraversion) and semantic overlaps among personality, emotion, and act representations persist. These correlations were treated as psychologically realistic priors rather than statistical nuisances.

3.3.1. Motivations and rationales of edge design

To jointly model emotion and dialogue act in conversations, we designed three complementary heterogeneous graphs (personality-aware utterance temporal graph [PAUTGAT], personality dual-task interaction temporal graph [PDITGAT], and personality-aware utterance and dual-task integrated temporal graph [PUDTGAT]) to capture dependencies among conversational units from different perspectives. The edge construction rules were not arbitrarily chosen; rather, they were motivated by well-established theories in dialogue linguistics and by the requirements of our joint tasks. Before formal mathematical exposition, we presented a unified conceptual framework to clarify the non-redundant, sequential roles of the three graph structures. **Figure 4** illustrates a minimal directed graph comprising three node types.

The three graphs (PAUTGAT, PDITGAT, PUDTGAT) implement sequential refinement stages rather than parallel redundant pathways. PAUTGAT establishes personality-conditioned semantic coherence; PDITGAT injects task-level interactions; PUDTGAT integrates both to produce the final prediction. This hierarchical decomposition ensures that each graph addresses distinct representational gaps.

- (i) PAUTGAT: modeling temporal structure and personality-aware utterance dependencies
 - (a) Motivation: Dialogue is inherently sequential and structured (e.g., turn-taking and adjacency pairs), and both dialogue acts and emotional states exhibit strong local dependencies across neighboring utterances. PAUTGAT was designed to encode such temporal discourse structure while conditioning the propagation of contextual signals on personality traits.

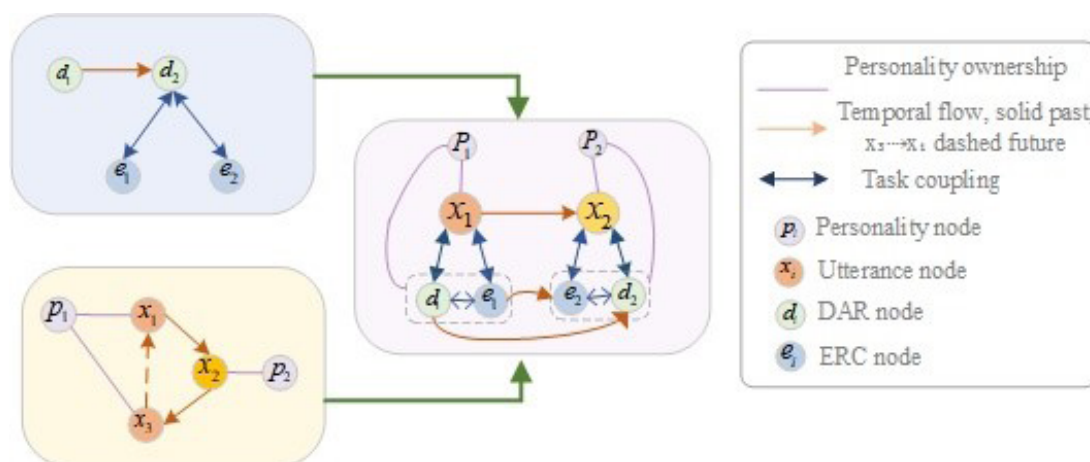


Figure 4. Conceptual logic graph: minimal node-edge structure

Abbreviations: DAR: Dialogue act recognition; ERC: Emotion recognition in conversation.

- (b) Edge design rationale: We primarily connected temporally adjacent utterances to capture the most fundamental local discourse relations (e.g., turn transitions and adjacency-pair-like patterns). When task characteristics suggest longer-range dependencies (e.g., topic initiation versus later summarization), additional edges were introduced between utterances that are far apart but structurally related. This principle is grounded in conversation analysis, especially studies of dialogue structure and adjacency pairs, which posit that crucial pragmatic dependencies often occur between neighboring turns while certain discourse functions span longer ranges.²⁵
- (ii) PDITGAT: modeling personality-conditioned dynamic influence and interaction
- (a) Motivation: Emotions in conversation are not isolated; they evolve through interaction and can be influenced by interlocutors. Importantly, the same semantic content may elicit different emotional/intentional reactions depending on the speaker's personality traits—i.e., personality is correlated with how affect is expressed and perceived.²⁶ PDITGAT aims to model dynamic influence and transition patterns within and across speakers under personality conditioning.
- (b) Edge design rationale: We introduced two main categories of temporal interaction edges:
- Intra-personality (intra-speaker) temporal edges: edges between adjacent utterances produced by the same speaker (equivalently, under the same personality profile), to capture the internal coherence and evolution of affective/intentional expression given a stable trait background.
 - Inter-personality (inter-speaker) interaction edges: edges between adjacent turns of different speakers, to model response and interaction patterns under differing personality traits (e.g., how a high-neuroticism speaker responds to a high-extraversion speaker).
- This design was inspired by the trait–situation interaction perspective in personality psychology and by pragmatic studies on individual speaking styles, suggesting that conversational responses may reflect both contextual stimuli and stable personal traits.²⁷
- (iii) PUDTGAT: modeling emotion–act synergy via cross-task coupling
- (a) Motivation: Emotion states and dialogue acts are pragmatically intertwined (e.g., a “challenge” act often co-occurs with “anger”, while “comfort” may co-occur with “sadness” or “neutral”). PUDTGAT was constructed to explicitly model such cross-task semantic coupling, enabling joint reasoning over emotion and act.
- (b) Edge design rationale: We added bidirectional edges between emotion and act nodes corresponding to the same utterance, enabling information exchange during encoding so that emotion prediction can benefit from act cues and vice versa. This is consistent with speech act theory and affective cognition views, which hold that dialogue acts function as carriers of affective expression.²⁸ We connected temporally adjacent emotion (and act) nodes to capture the continuity of a speaker's affective state, and connected adjacent-turn emotion nodes across speakers

to model interactive affect dynamics (e.g., alignment or escalation), which was motivated by the emotion contagion literature and studies on conversational emotion dynamics.²⁹

Overall, PAUTGAT emphasizes personality-aware utterance-level temporal structure, PDITGAT emphasizes personality-conditioned interaction and influence patterns, and PUDTGAT explicitly enforces emotion–act cross-task coupling. These graphs are complementary rather than redundant, collectively supporting joint inference through structurally grounded edge designs.

3.3.2. Personality-aware utterance temporal graph

We initially expounded on the construction of PAUTGAT, followed by an overview of the inferential computational processes it employs within the generated graph. To facilitate interactive information exchange among components, four distinct node types and four edge types have been devised.

We have established four types of representation nodes: personality nodes, utterance nodes, DAR nodes, and ERC nodes. Specifically, personality nodes represent the speaker's Big Five personality features, utterance nodes encapsulate discourse features, DAR nodes signify conversational act features within the dialogue, and ERC nodes denote emotional features within the discourse.

We constructed four types of edges: personality, utterance, ERC, and DAR. Among them, a personality edge connects nodes of discourse within the context of the same Big Five personality speaker. An utterance edge connects nodes in the context of discourse. DAR edge connects context act nodes that have mutual relationships

to simulate potential emotional dependencies between conversational acts. ERC edge connects context emotion nodes and dialogue acts to simulate explicit dependencies between two tasks.

Personality-aware utterance temporal graph is a comprehensive directed graph that associates a speaker's personality with discourse. It can simultaneously model more complete semantic dependencies within the same personality and between different personalities. Its definition is $G^{Paut} = (S^{Paut}, P^{Paut}, R^{Paut})$, where S^{Paut} represents the features of the utterance x_i and is initialized, P^{Paut} represents the adjacency matrix for the personality dependencies between utterance nodes and personality nodes, and R^{Paut} represents the adjacency matrix for the temporal dependencies between utterance nodes. A simplified schematic diagram containing only the smallest set of nodes and edges is shown in Figure 5.

We defined four types of dependency relationships: same-personality future, same-personality past, different-personality future, and different-personality past. Specifically, the same-personality future relationship refers to the relative position of the discourse under the same Big Five personality in the current conversation, being in the future. The same applies to the past relationship with the same personality. A different-personality future dependency relationship indicates a temporal relationship between the future and past positions of two discourses, even though they belong to different personalities. These four relationships simulate how past discourse influences future discourse, and vice versa, as well as the influence of discourse among different personalities. The detailed step-by-step computations are provided in Section A1.2.

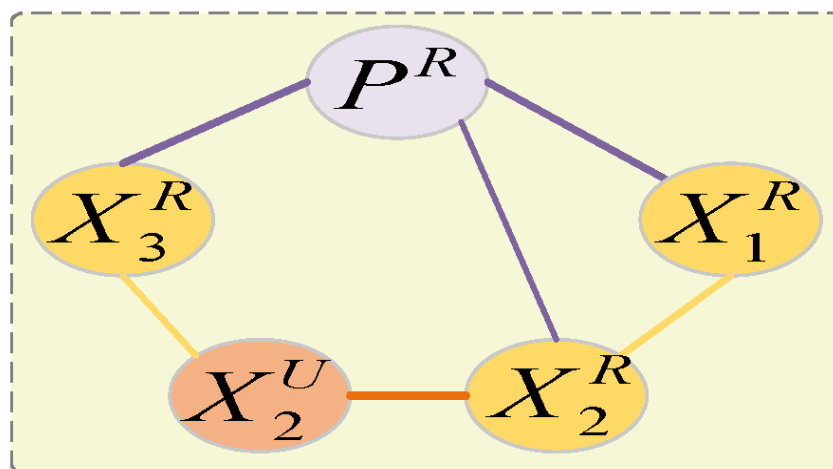


Figure 5. The simplest subgraph of a personality-aware utterance temporal graph. P^R denotes personality-aware representation; X_1^R , X_2^R , X_3^R denote role-aware utterance representation; X_2^U denotes utterance-level representation.

3.3.3. Personality dual-task interaction temporal graph

The personality dual-task interaction temporal graph is a comprehensive directed graph that associates the speaker's personality with dual tasks. It can simultaneously model more complete temporal interaction dependencies within tasks associated with personality and between tasks. PDITGAT uses DAR nodes, ERC nodes, ERC edges, and DAR edges to simulate interactions and propagation between tasks within and between different personalities. Similarly to PAUTGAT, we defined PDITGAT as $G^{Pdit} = \{(D^{Pdit}, P^{Pdit,D}, R^{Pdit,D}), (E^{Pdit}, P^{Pdit,E}, R^{Pdit,E})\}$, where D^{Pdit} represents the features of the act node d_i , $P^{Pdit,D}$ represents the adjacency matrix for the personality dependencies between act nodes and personality nodes, and $R^{Pdit,D}$ represents the adjacency matrix for the temporal dependencies between act nodes, $E^{Pdit}, P^{Pdit,E}, R^{Pdit,E}$ similarly. A simplified schematic diagram containing only the smallest set of nodes and edges is shown in **Figure 6**.

As in PAUTGAT, within each recognition task in PDITGAT, there are four relationship types. In the act task, we considered act information for the same personality in the past and future, as well as for different personalities in the past and future. In the emotion task, we considered emotion information for the same personality in the past and future, as well as for different personalities in the past and future. Simultaneously, we constructed connections across tasks in the timeline graphs to explicitly utilize interaction information. By doing so, we modeled these two sources of information within a unified graph interaction framework with inter-task connections. The detailed step-by-step computations are provided in **Section A1.3**.

3.3.4. Personality-aware utterance and dual-task integrated temporal graph

To further integrate information from the PAUTGAT and PDITGAT graphs, we expanded both graphs by utilizing directed edges in the PUDTGAT and applying specific rules to establish associations. This allowed us to focus on the interaction correlations between utterance and dual tasks under different personalities.

To further integrate information from the PAUTGAT and PDITGAT, the model employed mutual cross-attention as a bridge. The construction process of the PUDTGAT graph was similar to the PDITGAT graph construction method. After updating the nodes in PUDTGAT, it is represented as $D^{Puda} = \text{soft max}(X^{Paut} W_1 (D^{Pdit})^T)$, $E^{Puda} = \text{soft max}(X^{Paut} W_2 (E^{Pdit})^T)$.

To enhance learning from various graphs, multiple-head attention mechanisms were applied to graphs processed by the graph attention network (GAT) to further extract features. Additionally, to integrate the entire utterance and dual-task associations over time, the features from the GAT and the multi-head attention mechanism were fused globally using a gated recurrent unit (GRU). The detailed step-by-step computations are provided in **Section A1.4**.

3.4. Dual-task classification layer

After multiple interactions with the feature extraction initialization layer and the personality-fused interactive reasoning layer distributed and stacked with interaction layers, the model used co-attention to capture important relevant information. This process produced the final output of the interaction layer. Subsequently, separate decoders are employed for predicting conversational acts and emotions.

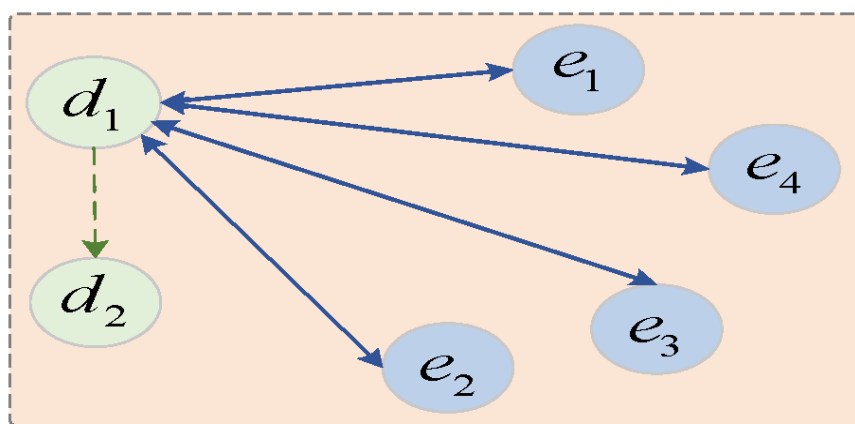


Figure 6. The simplest subgraph of the personality dual-task interaction temporal graph. d_i denotes discourse node representation; d_2 denotes previous dialogue state representation; e_1, e_2, e_3 , and e_4 denote emotion-related node representations.

The decoder consists of l identical blocks, where each block includes a multi-head self-attention layer, a co-attention layer, and a feedforward layer. In the emotion decoding process, we denoted the intermediate outputs of the self-attention sub-layer as $M_{E,part}^l$ and $M_{E,overall}^l$, the co-attention output as C_E^l , and the feed-forward output as E^l . Finally, the outputs of the last decoder block were fed into task-specific output layers to produce emotion and dialogue-act probability distributions at each time step. The detailed step-by-step computations of these sub-layer outputs are provided in Section A1.5.

3.5. Loss training

We adopted a multi-task learning framework to jointly minimize the emotion classification loss and the dialogue act recognition loss:

$$L_1 = -\sum_{i=1}^N \sum_{j=1}^{N_E} (\tilde{y}_i^{(j,E)} \log(y_i^{(j,E)}) + (1 - \tilde{y}_i^{(j,E)}) \log(1 - y_i^{(j,E)})) \quad (2)$$

$$L_2 = -\sum_{i=1}^N \sum_{j=1}^{N_D} (\tilde{y}_i^{(j,D)} \log(y_i^{(j,D)}) + (1 - \tilde{y}_i^{(j,D)}) \log(1 - y_i^{(j,D)})) \quad (3)$$

$$L = \frac{1}{2\sigma_E^2} L_E + \frac{1}{2\sigma_D^2} L_D + \log \sigma_E \sigma_D \quad (4)$$

where $\tilde{y}_i^E$ and $\tilde{y}_i^D$ are the true emotion labels and true act labels, respectively; N_E is the number of emotion labels, N_D is the number of act labels. σ_E and σ_D are the variances of emotion loss and act loss on the training instances.

In Equation 4, σ_E and σ_D are learnable task-weight parameters rather than fixed constants. They follow the homoscedastic-uncertainty weighting formulation for multi-task learning and were used to automatically balance the contributions of emotion classification and dialogue act recognition to the overall objective. Specifically, the terms $\frac{1}{2\sigma_E^2} L_E$ and $\frac{1}{2\sigma_D^2} L_D$ act as adaptive weights: during training, a task with higher intrinsic uncertainty (i.e., harder to learn) tends to be assigned a larger σ , which down-weights its loss and prevents it from dominating the gradient updates. The term $\log \sigma_E \sigma_D$ serves as a regularizer that constrains the scale of σ and avoids degenerate solutions where σ grows unbounded (i.e., the corresponding weight collapses to zero), improving identifiability and training stability. Consequently, we do not introduce additional manually tuned task-balancing coefficients; instead, σ_E and σ_D were optimized jointly with all other model parameters via backpropagation, enabling dynamic task balancing in an

end-to-end manner.

4. Experiments and analysis

4.1. Datasets

4.1.1. Dataset description

To comprehensively compare the performance of PEDR-GAT with existing classification models, we utilized three benchmark datasets spanning varying text lengths and multiple classification tasks.

- (i) Mastodon. It consists of 240 training, 266 testing, and 25 validation dialog sets. The dataset includes three emotion categories and 15 dialogue act categories.¹
- (ii) DailyDialog. It includes 11,118 training dialogues, 1,000 validation dialogues, and 1,000 test dialogues. The dataset includes seven emotion categories and four dialogue act categories.²⁴

Chinese psychological emotion dialogue contains 8,085 training dialogues, 933 validation dialogues, and 2,814 test dialogues. Each utterance has been annotated with Big Five personality traits. The dataset includes 13 emotion categories and 19 dialogue act categories.⁶

4.1.2. Personality annotation description

In addition to emotion and dialogue act annotations, the CPED dataset provides Big Five personality trait scores for each speaker. These scores were obtained from the standard self-report NEO five-factor inventory (NEO-FFI), which speakers complete independently outside the dialogue scenario. The NEO-FFI has been widely validated in psychological research, demonstrating strong psychometric quality—such as high internal consistency reliability (e.g., Cronbach's α) and test–retest reliability—thus providing a basic quality guarantee for the personality priors used in our study. The creators of CPED also report that the collected personality data exhibit good psychometric properties.

For DailyDialog and Mastodon—datasets lacking personality annotations—we used a BERT-based regression model trained on the essays and myPersonality datasets to predict Big Five trait scores from textual content. The resulting continuous scores were standardized and used as proxy personality embeddings, which were then fed into our PEDR-GAT architecture. The model achieved moderate concordance with self-reported traits (CCC = 0.42–0.51 across the five dimensions), consistent with prior work in computational personality prediction.³¹ We emphasize that these do not constitute validated psychological personality measures. Rather, they capture linguistic stylistic traces correlated with self-reported traits, including: lexical sophistication (word rarity, type-token ratio), syntactic

complexity (dependency depth, clause embedding), and discourse markers (hedges, intensifiers, self-references). Predicted embeddings reflect language styles typical of individuals who self-report high neuroticism/extraversion/etc., not necessarily the psychological mechanisms underlying those traits. Consequently, performance gains on DailyDialog and Mastodon should be interpreted as reflecting an indirect pathway: linguistic style → inferred personality profile → emotion/act prediction, whereas gains on CPED reflect direct personality modulation: self-reported trait → emotional/behavioral expression. To address potential information leakage concerns, we strictly ensured that the external data used for pretraining did not overlap with the downstream evaluation datasets. During PEDR-GAT training, these personality embeddings remain frozen to reduce overfitting and to maintain cross-task generalization.

In this work, we used each speaker's standardized scores on the five trait dimensions as static priors, assumed to be independent of the dialogue content, and fed them into our model as personality knowledge. We acknowledge that treating personality as static is a simplifying assumption; its limitations will be discussed in Section 5.2.1.

4.2. Competitor methods

In dialogue emotion recognition, a sequence-based comparative model is DialogueRNN, a recurrent neural network model.

In dialogue emotion recognition, a graph neural network-based comparative model is DialogueGCN, a speaker-aware model.

In dialogue act recognition, two sequence-based comparative models are:

- Hierarchical emotion classifier (HEC)¹²: A dialogue act recognition model based on a bidirectional time-series structure.
- Context-aware sentiment analysis (CASA)¹³: A dialogue act recognition model based on sequence fusion with a self-attention mechanism.
- In joint dialogue emotion classification and act recognition, four sequence-based comparative models are:
- Joint dialogue act and sentiment recognition (JointDAS)¹: A joint task recognition model based on the RNN structure for joint dialogue emotion classification and act recognition.
- Interactive information integration model (IIIM)²: A model based on a convolutional neural network (CNN) that integrates act, emotion, and predictor for joint dialogue emotion classification and act recognition tasks.

- Deep collaborative relationship network (DCR-Net)³: A model that utilizes a deep collaborative interaction relationship network to explicitly model the influence between act and emotion classification tasks for joint recognition.
- Bi-channel dialogue context network (BCDCN)¹⁷: A context-aware dynamic convolutional network extended to a dual-channel version for joint dialogue emotion classification and act recognition tasks.
- In joint dialogue emotion classification and act recognition, three graph neural network-based comparative models are:
- Conversational graph attention network (Co-GAT)²²: A joint interactional graph attention network model with cross-dialogue connections for both dialogue emotion classification and act recognition tasks.
- Dialogue act and emotion recognition (DARER)⁵: A model based on LSTM and utilizing graph structure for temporal representation in the discourse task, predicting hierarchical relationships with task temporal information for joint dialogue emotion classification and act recognition tasks.
- Task-specific contrastive learning (TSCL)⁷: A joint dialogue emotion classification and act recognition model based on bidirectional LSTM and incorporating speaker graph structure to transform embedded discourse into confidence scores.

4.3. Experiment settings

We implemented all models in PyTorch²⁵ and trained them on an NVIDIA RTX 3060 Super GPU. The hyperparameters for PEDR-GAT were set as follows: embedding dim = 64, hidden dim = 256, dropout = 0.3, number of layers = 1, batch size = 128, and maximum sequence length = 64. We chose AdamW as the optimizer with a learning rate of 0.0001. Each model was trained for 50 epochs. To ensure computational transparency, we reported comprehensive cost metrics in Table 2. To disentangle architectural design from parameter stacking effects, we constructed a capacity-matched control model (PEDR-GAT-C) that matches PEDR-GAT's total parameters (6.3M) by: (a) increasing GAT hidden dimensions from 256 to 320; (b) adding two extra GAT layers; while ablating personality nodes and cross-task edges.

To mitigate the risk of overfitting introduced by the architecture (three GAT networks combined with multi-head attention and a GRU), we additionally adopted explicit regularization and training strategies beyond the hyperparameter settings above. Specifically, we applied dropout after every GAT layer output (dropout rate = 0.5) to reduce feature co-adaptation, and incorporated L_2 regularization via weight decay ($weight\ decay = 1 \times 10^{-4}$) in

Table 2. Computational cost comparison

Model	Parameters (M)	FLOPs (G)	GPU memory (GB)	Epoch (min)	Batch (ms)
DialogueGCN	3.2	12.5	4.8	8.5	45
Co-GAT	5.8	28.3	6.2	12.3	62
DARER	6.1	31.7	7.1	14.6	78
PEDR-GAT	6.3	34.2	7.5	16.2	85
PEDR-GAT-C	6.3	33.8	7.4	15.8	82

Abbreviations: Co-GAT: Conversational graph attention network; DARER: Dialogue act and emotion recognition; DialogueGCN: Dialogue graph convolutional network; FLOP: Floating point operations per second; GPU: Graphics processing unit; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition; PEDR-GAT-C: Personalized model integrating personality for joint dialogue emotion classification and act recognition with context modeling.

the optimizer to constrain model complexity. We further employed early stopping by monitoring the validation joint loss: training was terminated if the validation loss did not improve for 10 consecutive epochs, and we restored/saved the checkpoint with the lowest validation loss for final evaluation. These measures help improve generalization and prevent overfitting.

4.4. Experiment I: Validation experiment

Experiment I compared the accuracy of PEDR-GAT with that of baseline models on the Mastodon, DailyDialog, and CPED datasets. We compared PEDR-GAT with emotion recognition, act recognition, and joint-learning models on both the Mastodon and DailyDialog datasets. From the experimental results (Tables 3 and 4), we can draw the following conclusions:

- (i) The joint model performs better overall than individual recognition models on both datasets, demonstrating the potential connection and mutual promotion between these two tasks.
- (ii) Our model achieves comparable or better results on most evaluation metrics for both datasets. Specifically, on the Mastodon dataset, the F1 metric for the ERC task is 1.9% higher than that of the latest joint model, DARER, while the F1 metric for the DAR task decreases slightly by 0.7%. On the DailyDialog dataset, the F1 metric for the ERC task is 2.0% higher than the latest joint model DARER, while the F1 metric for the DAR task slightly decreases by 0.6%. On the DailyDialog dataset, the F1 metric for the ERC task increases by 2.3%, and the F1 metric for the DAR task increases by 1.9%. The results for act recognition in Mastodon show a slight decrease, which may be attributed to imbalanced act annotations in the dataset.

The model's ability to achieve better results stems from (i) the model not only considers the discourse structure related to different speaker personalities but also takes

into account the interactive influences within and between tasks under personality fusion. (ii) Through local utterance learning and dual-task interactive learning with utterance context, the model can comprehensively capture the feature information of both tasks in the dialogue. (iii) Edge categories in the graph, along with the relative position vectors of different nodes, were utilized to update edge weights, considering the speaker's emotional dependency, act dependency, and temporal features.

Compared to the two datasets, the CPED dataset has already been annotated with Big Five personality traits for dialogue sentences. The model can more accurately learn the accompanying personality features of the sentences, resulting in superior recognition results, as shown in Table 5.

The experimental results indicate that, for joint tasks, PEDR-GAT outperforms the best baseline (DARER), achieving a 5.3% increase in F1 for the ERC task and a 2.3% increase for the DAR task.

A personalized, personality-integrated model is evaluated for joint dialogue emotion classification and act recognition on the CPED test set, with validation at two granularity levels: emotion category and behavior category. The accuracy heatmap of PEDR-GAT is shown in Figure 7.

Figure 7 presents the confusion matrix for emotion classification for the full model on the CPED test set. Overall, the model exhibits systematic rather than random errors, as evidenced by salient confusions between sadness and fear and between anger and disgust, which is consistent with their similarity along affective dimensions such as valence and arousal. In contrast, the proportion of happiness misclassified as neutral is relatively low, suggesting that the model captures positive affect cues more reliably. Beyond these general patterns, the impact of personality is emotion-specific rather than uniformly improving across all categories.

Table 3. Comparative experiments on the CPED dataset

Model	ERC			DAR		
	Precision	Recall	F1	Precision	Recall	F1
DialogueRNN	0.405	0.428	0.416	–	–	–
DialogueGCN	0.414	0.434	0.423	–	–	–
HEC	–	–	–	0.565	0.557	0.5600
CASA	–	–	–	0.557	0.571	0.563
JointDAS	0.361	0.416	0.386	0.556	0.516	0.535
IIIM	0.387	0.401	0.393	0.563	0.522	0.541
DCR-Net	0.432	0.473	0.451	0.603	0.569	0.585
BCDCN	0.382	0.620	0.472	0.573	0.617	0.594
Co-GAT	0.440	0.532	0.481	0.604	0.606	0.604
TSCL	0.461	0.587	0.516	0.612	0.616	0.613
DARER	0.560	0.633	0.594	0.650	0.618	0.633
PEDR-GAT	0.583	0.649	0.614	0.643	0.612	0.627

Abbreviations: BCDCN: Bi-channel dialogue context network; CASA: Context-aware sentiment analysis; Co-GAT: Conversational graph attention network; DAR: Dialogue act recognition; DARER: Dialogue act recognition and emotion recognition; DCR-Net: Deep collaborative relationship network; DialogueGCN: Dialogue graph convolutional network; DialogueRNN: Dialogue recurrent neural network; ERC: Emotion recognition in conversation; HEC: Hierarchical emotion classifier; IIIM: Interactive information integration model; JointDAS: Joint dialogue act and sentiment recognition; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition; TSCL: Task-specific contrastive learning.

Table 4. Comparative experiments on the DailyDialog dataset

Model	ERC			DAR		
	Precision	Recall	F1	Precision	Recall	F1
DialogueRNN	0.445	0.377	0.408	–	–	–
DialogueGCN	0.418	0.445	0.431	–	–	–
HEC	–	–	–	0.778	0.765	0.771
CASA	–	–	–	0.779	0.765	0.771
JointDAS	0.354	0.288	0.317	0.762	0.745	0.753
IIIM	0.389	0.285	0.328	0.765	0.749	0.756
DCR-Net	0.560	0.401	0.467	0.791	0.790	0.790
BCDCN	0.552	0.457	0.500	0.800	0.806	0.802
Co-GAT	0.659	0.453	0.536	0.810	0.781	0.795
TSCL	0.566	0.492	0.526	0.788	0.798	0.792
DARER	0.599	0.495	0.542	0.813	0.808	0.810
PEDR-GAT	0.623	0.517	0.565	0.834	0.826	0.829

Abbreviations: BCDCN: Bi-channel dialogue context network; CASA: Context-aware sentiment analysis; Co-GAT: Conversational graph attention network; DAR: Dialogue act recognition; DARER: Dialogue act recognition and emotion recognition; DCR-Net: Deep collaborative relationship network; DialogueGCN: Dialogue graph convolutional network; DialogueRNN: Dialogue recurrent neural network; ERC: Emotion recognition in conversation; HEC: Hierarchical emotion classifier; IIIM: Interactive information integration model; JointDAS: Joint dialogue act and sentiment recognition; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition; TSCL: Task-specific contrastive learning.

Table 5. Comparative experiments on the CPED dataset

Model	ERC			DAR		
	Precision	Recall	F1	Precision	Recall	F1
JointDAS	0.323	0.263	0.289	0.536	0.476	0.504
IIIM	0.344	0.259	0.295	0.542	0.483	0.510
DCR-Net	0.511	0.372	0.430	0.580	0.529	0.553
BCDCN	0.504	0.398	0.444	0.555	0.582	0.568
Co-GAT	0.595	0.386	0.468	0.575	0.577	0.575
TSCL	0.513	0.422	0.463	0.584	0.585	0.584
DARER	0.556	0.447	0.495	0.604	0.587	0.595
PEDR-GAT	0.647	0.476	0.548	0.637	0.601	0.618

Abbreviations: BCDCN: Bi-channel dialogue context network; Co-GAT: Conversational graph attention network; DAR: Dialogue act recognition; DARER: Dialogue act recognition and emotion recognition; DCR-Net: Deep collaborative relationship network; ERC: Emotion recognition in conversation; IIIM: Interactive information integration model; JointDAS: Joint dialogue act and sentiment recognition; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition; TSCL: Task-specific contrastive learning.



Figure 7. Heatmap of effectiveness experiment. (A) Accuracy of emotion category perception. (B) Accuracy of action category perception.

This observation is consistent with the category-level analysis of perception accuracy. PEDR-GAT achieves perception accuracies above 0.6 for happy, relaxed, astonished, worried, and negative-other, while the remaining emotion categories are all above 0.4—well above the random perception probability of 0.2. The lowest accuracy is observed for depression (0.41). Notably, negative emotion categories exhibit higher perception accuracy than positive ones overall, suggesting that the joint model is particularly sensitive to negative affect signals in CPED. For dialogue acts, PEDR-GAT attains perception accuracies above 0.5 for question, command, apology, thanking, disagreement, and greeting; comfort is the lowest but still above the random baseline. All act categories achieve accuracies above 0.3. Although class imbalance in dialogue acts limits the overall accuracy, act cues still provide complementary information and can interact with emotion categories to facilitate joint reasoning.

To validate the mechanistic roles of specific graph structures rather than redundant design, we conducted a triangulated analysis linking observed confusion patterns, speaker personality profiles, and architectural component contributions, as shown in Table 6.

Pattern A: sadness ↔ fear. This confusion is concentrated among speakers with high neuroticism (>0.7) and low conscientiousness (<0.4)—the “vulnerable instability” profile. PDITGAT removal increases this confusion by 12.4%, whereas PAUTGAT (+2.1%) and PUDAGAT (+1.7%) have minimal impact. High-neuroticism speakers conflate present sorrow with prospective distress; PDITGAT resolves this by coupling emotions (sadness → comfort-seeking versus fear → information-seeking). Pattern B: anger ↔ disgust. Low-Agreeableness speakers (<0.3) account for 65.7% of these errors. PAUTGAT removal increases confusion by 9.8%, implicating its personality-anchored discourse history. Low-agreeableness speakers develop idiosyncratic linguistic signatures (e.g., sarcastic disgust versus direct anger); PAUTGAT leverages trait-stabilized contextual priors to disambiguate. This analysis reveals functional specialization in architectural complexity: each graph targets personality-specific

confusion patterns. Marginal F1 gains aggregate targeted error reductions that are invisible at the macro level yet critical for deployment.

Overall, these results suggest that personality information does not yield uniform gains across all categories. Instead, it selectively strengthens the model’s ability to differentiate semantically and pragmatically similar emotions—particularly negative and complex emotions whose interpretation depends on speaker-specific tendencies—while dialogue acts provide additional cues that promote joint reasoning despite label imbalance. We attribute the improvements to three complementary design choices: (i) bidirectional temporal modeling of speakers’ Big Five traits in dialogue contexts, (ii) the personality-associated dual-task interaction graph that explicitly models intra-task dependencies and inter-task correlations under personality conditioning, and (iii) the feature extraction initialization layer, which leverages label information to form personality-guided emotion and act representations and enrich contextual features.

4.5. Experiment II: Ablation study

4.5.1. Ablation settings and result analysis

To further validate the effectiveness of each module in PEDR-GAT, experiment II conducted ablation studies on CPED and DailyDialog. Ablation studies typically involve removing specific features of a model or algorithm and observing how it affects the model’s performance. The experiments mainly involved removing model features individually, including removing the personality-label information and the initial estimates in the feature extraction initialization layer, removing the models corresponding to PDITGAT, PAUTGAT, and PUDAGAT in the personality-fused interactive reasoning layer, and removing the model with co-attention in the dual-task classification layer. The F1 scores on the datasets were then calculated, as shown in Figure 8.

After removing personality nodes, the F1 scores for anger and sadness drop the most (both by more than 5 points), and their confusion with neutral increases substantially. This is consistent with the psychological expectation that speakers

Table 6. Confusion–personality–component triangulation

Confusion patterns	Personality profile	Critical component	Mechanism
Sadness ↔ fear	High neuroticism + low conscientiousness	PDITGAT	Cross-task emotion-act coupling
Anger ↔ disgust	Low agreeableness	PAUTGAT	Trait-anchored discourse history
Happiness → neutral	High extraversion + high agreeableness	PUDAGAT	Global personality-context fusion

Abbreviations: PAUTGAT: Personality-aware utterance temporal graph; PDITGAT: personality dual-task interaction temporal graph; PUDTGTAT: Personality-aware utterance and dual-task integrated temporal graph.

high in neuroticism tend to experience and express negative affect more frequently or more intensely; personality cues help the model disambiguate subtle negative expressions that would otherwise collapse into neutral expressions. In contrast, happiness remains relatively stable under ablation (an F1 decrease of approximately 2.9 points) and shows limited confusion with neutral, plausibly because happiness is often expressed through overt lexical and

stylistic markers. Nevertheless, personality information—particularly extraversion—can still reduce residual confusions with other high-arousal positive emotions such as surprise. Finally, personality cues also enhance fine-grained discrimination between closely related negative emotions, especially the frequently confused pair anger versus disgust. A plausible explanation is that speakers with different traits (e.g., conscientiousness) may adopt

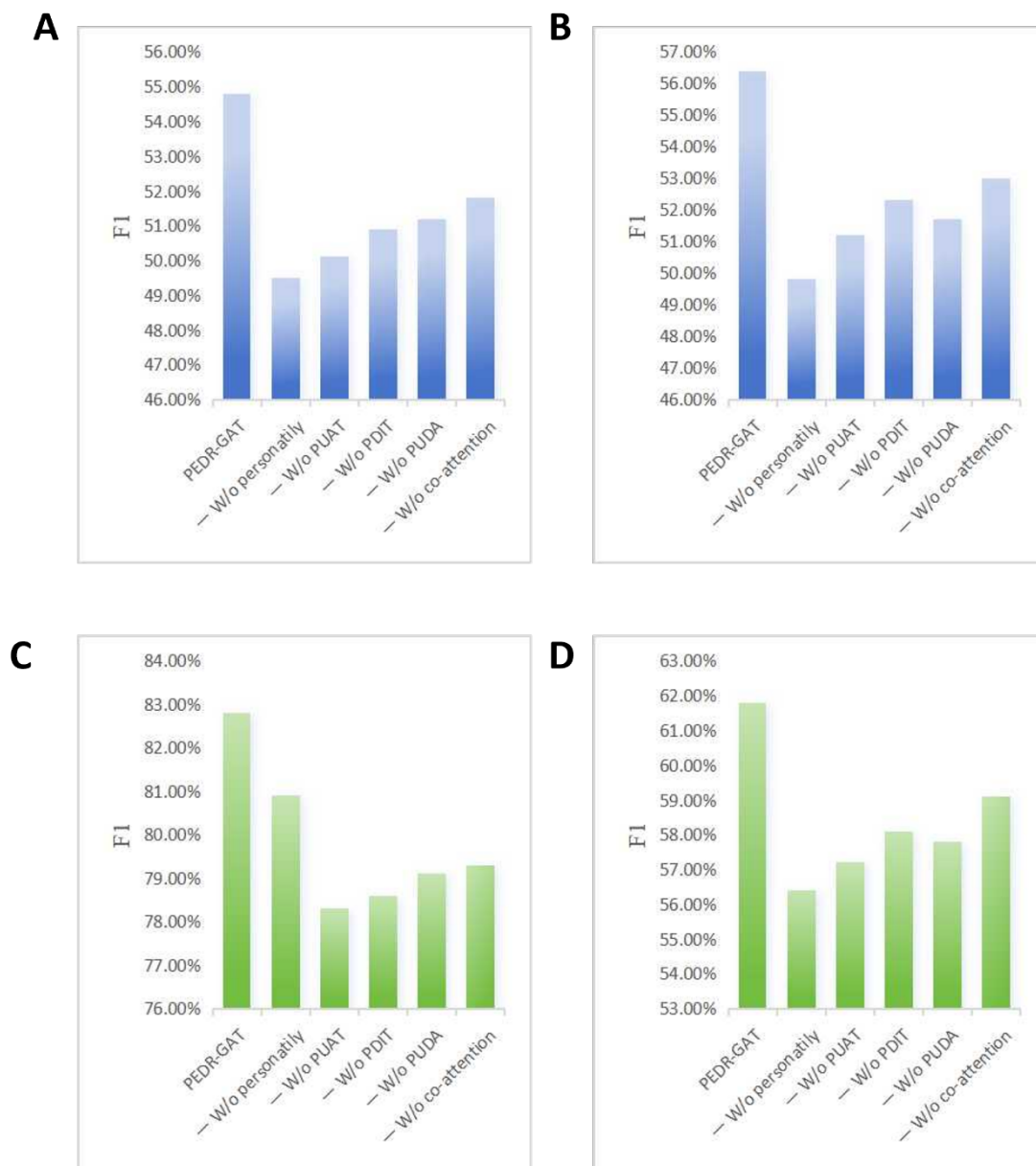


Figure 8. Results of ablation experiments. (A) Accuracy of emotion category on DailyDialog. (B) Accuracy of the action category on DailyDialog. (C) Accuracy of emotion category on CPED. (D) Accuracy of the action category on CPED.

Abbreviations: CPED: Chinese psychological emotion dialogue; PDIT: Personality-aware dual-task interaction; PEDR-GAT: Personality-enhanced dialogue representation graph attention network; PUAT: Personality-aware utterance attention; PUDA: Personality-aware utterance-level dual-task attention.

different lexical choices and pragmatic strategies when expressing criticism or dissatisfaction; personality nodes allow the model to capture these trait-conditioned patterns and improve separation among similar classes.

Figure 9 illustrates the relationship between parameter count and validation performance across different model variants. As can be observed, the full model (PEDR-GAT) achieves a substantial performance jump with only a modest increase in parameters, whereas baseline models that merely increase model depth (e.g., by adding more GAT layers) exhibit diminishing returns. This further indicates that our gains primarily stem from the effectiveness of the architectural design, rather than simple parameter stacking.

From the results of the ablation experiments, we can draw the following conclusions:

- (i) Removing personality features diminishes the model's ability to learn the dependency between speakers and both types of labels, resulting in weakened correlation between emotions and behaviors, leading to decreased

performance.

- (ii) After removing the personality-aware utterance attention (PUAT), personality-aware dual-task interaction (PDIT), and personality-aware utterance-level dual-task attention (PUDA) graph structures, PEDR-GAT struggles to capture temporal information between utterances, temporal interaction between tasks, and establish correlations between utterances and tasks. PUAT, as the semantic understanding core anchored by personality, captures semantic transitions under different personalities, providing basic context for perception tasks. Removing PDIT significantly reduces PEDR-GAT's performance, impairing mutual prompting inference between the internal transformations and external interactions of the two tasks. PUDA integrates semantics and dual-task relationships above personality. Removing these graph structures reduced accuracy, demonstrating that personality, semantic, and temporal correlations contribute to recognition tasks.
- (iii) In the decoding stage, co-attention integrates the

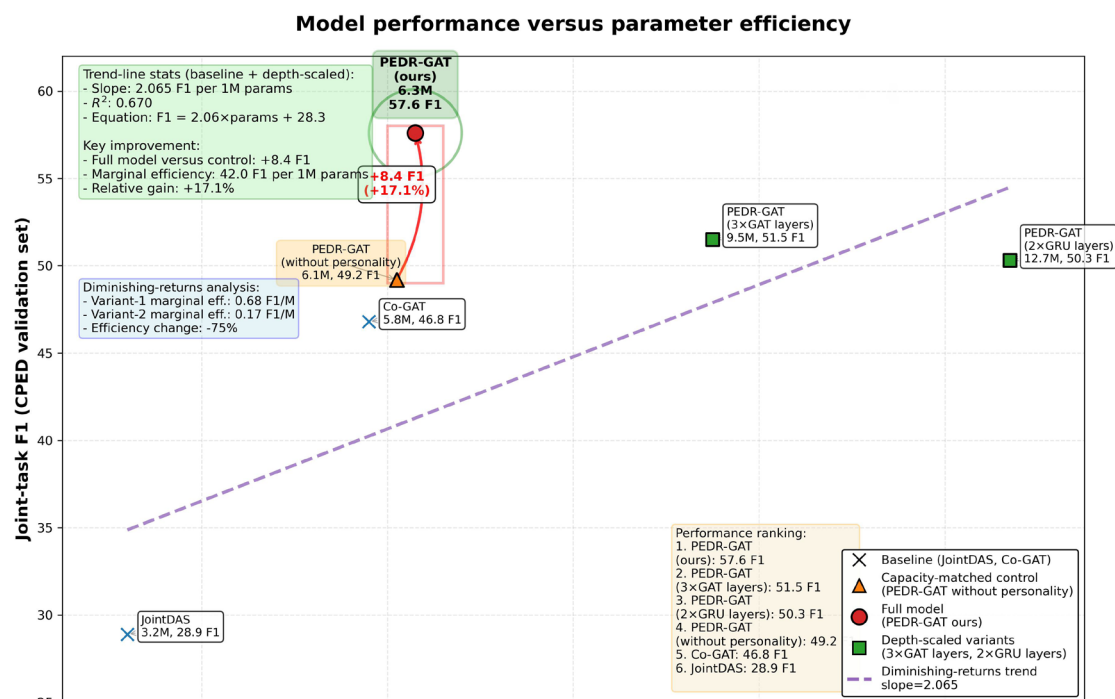


Figure 9. Model performance and parameter–efficiency analysis. The figure shows the relationship between the joint-task F1 score and the parameter count of different model variants on the CPED validation set. Key findings: (i) PEDR-GAT (○) and the capacity-matched control model-PEDR-GAT-C (△) have comparable parameter counts, but PEDR-GAT achieves substantially higher performance, demonstrating that the gains arise from the personality-aware design rather than increased capacity; (ii) the diminishing-returns trend line (blue dashed line) indicates that simply increasing parameters is inefficient; (iii) the full model surpasses the efficiency boundary and achieves superior parameter efficiency. Abbreviations: Co-GAT: Conversational graph attention network; CPED: Chinese psychological emotion dialogue; GAT: Graph attention network; GRU: Gated recurrent unit; JointDAS: Joint dialogue act and sentiment recognition; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition; PEDR-GAT-C: Personalized model integrating personality for joint dialogue emotion classification and act recognition with context modeling.

initial dual-task recognition features fused with the Big Five personality information and the globally integrated dual-task recognition features fused with semantics through dual-task interaction. The model better considers dual-task-related information in both local and global contexts, accounting for personality involvement through a shared attention mechanism.

To ensure the robustness of our results, we reported the average performance and standard deviation over 10 independent runs with different random seeds. As shown in Table 7, statistical significance between models is assessed using a paired permutation test with 10,000 permutations and a significance level of $\alpha = 0.05$.

Table 7. Robustness evaluation—mean and standard over 10 runs (different random seeds)

Model	DailyDialog (F1-emo)	CPED (F1-emo)
DARER	55.4	61.2
Our model	58.6 ^a	64.8 ^a

Notes: ^aThe best results. Results marked with * are statistically significantly better than all baselines ($p < 0.05$). Abbreviations: CPED: Chinese psychological emotion dialogue; DARER: Dialogue act recognition and emotion recognition.

Improvements over the strongest baseline (dialogue graph-based emotion recognition with edge relations) are statistically significant ($p < 0.01$) on DailyDialog and CPED by a paired permutation test (Section A4). Our proposed PEDR-GAT consistently outperforms baseline models across all experimental datasets. While absolute F1 improvements may appear modest, their practical significance becomes clearer when considering error severity rather than aggregate accuracy. Even 1% gains in emotion recognition reduce user frustration in human-computer interaction—for example, detecting neutral-to-

anger shifts can prevent escalation in customer service. More critically, our model achieves approximately a 15% reduction in error severity in the distress→happiness and suicidal cue omission analyses (Section A3). These rare but severe failures carry disproportionate consequences; their reduction validates marginal F1 gains as meaningful safety improvements in deployment.

The ablation study confirms that each component contributes positively to the final performance. The performance drop upon removing the personality nodes is not only substantial in magnitude but also statistically significant ($p < 0.001$), underscoring the critical role of personality infusion.

4.5.2. Fine-grained error analysis on chinese psychological emotion dialogue

To further investigate the practical contribution of the personality-aware module in Chinese conversational scenarios, we conduct a fine-grained error analysis of the CPED dataset using an ablation setting that removes the personality nodes. Table 8 reports the performance changes of PEDR-GAT on specific emotion and dialogue act categories in CPED after removing personality nodes.

Based on the above results, the following analysis was conducted:

- Pronounced impact on high-arousal and negative emotions. On CPED, the largest performance drops are observed for sadness (−5.3), anger (−5.1), and the disagreement act (−5.2). This phenomenon is closely related to the implicit nature of emotional expression in Chinese conversations and to the moderating role of personality traits. For example, individuals with higher neuroticism may be more likely to internalize negative affect, resulting in more subtle expressions of “sadness.” Without personality information,

Table 8. Fine-grained error analysis on CPED

	PEDR-GAT(F1)	Ablation model (F1)	$\Delta F1$	Main confusion direction
Emotion categories				
Happiness	60.6	57.2	−3.4	Neutral, surprised
Anger	57.4	52.3	−5.1	Disgusting, neutral
Sadness	56.1	50.8	−5.3	Neutral, worried
Dialogue act categories				
Question	70.3	67.4	−2.9	Comfort, statement-opinion
Statement-opinion	68.4	64.2	−4.2	Disagreement, response
Disagreement	65.3	60.1	−5.2	Statement, questioning

Abbreviation: CPED: Chinese psychological emotion dialogue; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition.

the model becomes less able to capture individual differences and is more likely to misclassify these utterances as neutral. As illustrated in the example (an ironic disagreement utterance produced by a speaker with high openness), the full model can leverage the speaker's personality profile (e.g., being more open-minded and verbally expressive) to correctly recognize both the disagreement act and the anger emotion, whereas the ablated model tends to predict a neutral statement.

- Critical role in recognizing the “expressing emotion” act, which is common in Chinese dialogue. The CPED-specific dialogue act category expressing emotion is substantially affected, with an F1 decrease of 4.7 points. This act is highly dependent on individual tendencies: speakers with higher extraversion may express their emotions more directly, while those with higher conscientiousness may express them more restrainedly. After removing personality cues, the model has difficulty distinguishing explicit “expressing emotion” utterances from factual Statement-like utterances.
- Personality effects on face-threatening acts in Chinese conversations. In Chinese interaction, personality traits such as agreeableness may be linked to systematic variation in how face-threatening acts (e.g., disagreement, criticism) are expressed. Our analysis shows that for speakers with lower agreeableness who express disagreement more directly, the ablated model is more prone to confusion and misclassification. This suggests that personality information plays an important role in understanding the interaction between culture-specific pragmatic norms and individual speaking styles.

4.5.3. Stratified analysis by personality contrast

To validate whether the observed performance gains genuinely stem from personality-informed modeling rather than contextual confounds, we conducted a stratified analysis based on the personality contrast. On CPED,

where ground-truth personality scores were available, we computed the standard deviation of each Big Five dimension across speakers in a dialogue, then averaged across five dimensions. Dialogues with mean $\sigma > 0.8$ (above median) constituted the high-contrast group; those with mean $\sigma \leq 0.8$ formed the homogeneous-profile group. The rationale is that conversations with larger personality variance across speakers provide more discriminative personality signals, whereas homogeneous conversations offer weaker personality cues.

Based on the above results, the following analysis was conducted:

- Performance gains concentrated in high-contrast group: The full PEDR-GAT achieves substantially larger F1 improvements in high-contrast conversations (+5.6 ERC-F1, +3.5 DAR-F1) compared to low-reliability conversations (+0.6 ERC-F1, +0.5 DAR-F1). This pattern demonstrates that personality information contributes discriminatively when personality signals are strong, but provides marginal benefit when speakers are psychologically homogeneous.
- Diminishing returns under weak signals: In the homogeneous-profile group, removing personality nodes results in minimal performance degradation, suggesting that the model effectively falls back to context-only reasoning when personality cues are unreliable. This validates the architecture's robustness but also underscores that “pure” personality effects are contingent on annotation quality.
- Cross-group generalization: The high-contrast group's gains (+10.3 ERC-F1 over the low-reliability baseline) substantially exceed the sum of the personality effects, indicating a synergistic interaction between strong personality signals and contextual temporal modeling. This supports our core claim that personality-guided temporal graphs capture trait-conditioned interaction patterns inaccessible to context-only models.

The stratified results caution against over-interpreting F1 gains on DailyDialog and Mastodon. Since predicted

Table 9. PEDR-GAT high versus low-reliability personality group comparison

Model variant	High-contrast group ($\sigma > 0.8$)	Homogeneous-profile group ($\sigma \leq 0.8$)	Δ (High–low)			
	ERC-F1	DAR-F1	ERC-F1	DAR-F1	ERC	DAR
PEDR-GAT (Full)	62.4	64.7	52.1	60.3	+10.3	+4.4
Without personality	56.8	61.2	51.5	59.8	+5.3	+1.4
Gain from personality	+5.6	+3.5	+0.6	+0.5		

Abbreviation: DAR: Dialogue act recognition; ERC: Emotion recognition in conversation; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition.

personality embeddings likely exhibit homogeneous, low-variance distributions (regression models tend toward mean predictions), the effective “reliability” of these pseudo-personality signals is low. Thus, the observed improvements on these datasets (+1.9% to +3.3%) probably reflect linguistic style exploitation (high-contrast-like gains from stylistic distinctiveness) rather than genuine personality modulation. Future work should explicitly model prediction uncertainty to dynamically weight personality influence.

4.6. Experiment III: Case study

To qualitatively examine how personality-conditioned temporal reasoning shapes joint emotion and dialogue-act predictions, we presented two case studies (Tables 10 and 11).

In both test cases, when speakers A and C share the same Big Five personality traits, they exhibit highly consistent matching of behavior and emotion despite facing different personalities. From these examples, we observe that emotions are continuous over time across different conversational contexts, while behaviors are temporally linked, consistent with the logic of discourse time reasoning and dual-task time reasoning in PEDRGAT. However, inconsistencies may arise between the recognized emotions and the emotions expressed in the statements, possibly due to limited annotations of Big Five personality traits in the training data and data imbalance.

4.7. Experiment IV: Robustness to noisy and sparse personality signals on conversational personality–emotion dataset

To evaluate the model’s sensitivity to imperfect personality information, we conducted additional robustness experiments on CPED:

- (i) Annotation noise. We injected uniform noise into personality labels by randomly perturbing each trait by ± 1 Likert level. The results indicate that the model is most sensitive to noise in extraversion and neuroticism, which are also the two traits most strongly associated with emotion recognition performance.
- (ii) Label sparsification. To simulate realistic scenarios where only a subset of users have personality profiles, we randomly masked personality labels for 30% and 50% of speakers by setting their personality vectors to zero (treated as unknown). As shown in Table 12, the performance drop suggests that the model can partially compensate via dialogue context, yet missing personality information still causes a noticeable degradation—particularly for happiness and sadness. Based on our findings on CPED, we discuss several

directions toward practical deployment:

- Context-aware dynamic personality modeling. Future work may explore lightweight adapters that refine base personality traits in context, based on dialogue topics and interlocutor feedback, making personality representations more compatible with the dynamics of Chinese interaction.
- Jointly inferring personality from text. For real-world applications without explicit personality annotations, one can design pretraining objectives for large-scale Chinese dialogues to predict language-style cues, affective lexical usage patterns, and other personality-related surface features as proxy signals for personality traits.
- Building cross-cultural personality dialogue benchmarks. We encourage the community to develop Chinese dialogue datasets with finer-grained personality expressions and multi-scenario behaviors, to better evaluate and improve generalization in practical settings.

Overall, our in-depth analysis of CPED confirms that personality information is crucial for understanding emotions and dialogue acts in Chinese conversations, especially for distinguishing implicit expressions in high-context cultures. At the same time, we acknowledge the limitations of modeling personality as static self-reported labels. The robustness experiments demonstrate that the model has a certain degree of resilience to annotation imperfections; however, moving toward real-world deployment will require advancing personality modeling toward more dynamic representations that are easier to infer directly from data.

5. Discussion

5.1. Model overhead

In the inference layer, this model employs three graph attention networks to characterize the temporal interaction structures among personality–utterance, personality–dual tasks, and utterance–dual tasks (PAUTGAT, PDITGAT, PUDTGAT). After the graph updates, multi-head attention and a GRU are further introduced for global fusion. Meanwhile, in the classification layer, the model employs a decoder stack comprising multi-head self-attention, co-attention, and feed-forward layers to perform dual-task prediction. Overall computational overhead is primarily determined by the multi-graph GAT updates and the attention mechanisms during the decoding stage.

5.1.1. Notation and graph scale

Assume the dialogue history contains N utterances and S speakers (or personality nodes). Let the hidden dimension

Table 10. PEDR-GAT personality-aware joint task recognition test case 1

Speaker's Big Five personality	Utterance	Action recognition objective	Action recognition outcomes	Emotion recognition objective	Emotion recognition outcomes
A: [High high low high low]	I think he would be more nervous about getting her a shrink.	Disagreement	Statement-opinion	Worried	Worried
B: [Low high low low high]	We need to find a professional doctor.	Statement-opinion	Statement-opinion	Neutral	Sadness
A: [High high low high low]	I am just suggesting that we plan to do it, not rush it.	Statement-opinion	Disagreement	Anger	Worried
A: [High high low high low]	I think psychiatrists nowadays are as unprofessional as those so-called counseling companies.	Statement-opinion	Statement-opinion	Disgust	Disgust
B: [Low high low low high]	You have not seen a psychiatrist. How do you know?	Question	Question	Neutral	Neutral

Abbreviation: PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition.

Table 11. PEDR-GAT personality-aware joint task recognition test case 2

Speaker's Big Five personality	Utterance	Action recognition objective	Action recognition outcomes	Emotion recognition objective	Emotion recognition outcomes
C: [High high low high low]	You drew a cat's head in six days.	Question	Question	Astonished	Astonished
D: [Low high low low high]	I cannot find it in me to paint.	Answer	Statement-opinion	Depressed	Sadness
C: [High high low high low]	If you are in a bottleneck, you need to relax, or I will go for a walk with you.	Question	Question	Neutral	Worried
C: [High high low high low]	I just need to go to the supermarket.	Question	Comfort	Neutral	Neutral
D: [Low high low low high]	I will get you anything you want.	Statement-opinion	Statement-opinion	Neutral	Neutral

Abbreviation: PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition.

Table 12. Robustness analysis of personality noise/sparsity on Chinese personality and emotion dialogue

Experimental setting	Emotion (F1)	Dct (F1)	$\Delta F1$ (Emo)	Most affected emotion categories
Full personality labels	54.8	61.8	–	–
+ Uniform noise (± 1 Likert level)	53.4	60.2	–1.4	Anger, happiness
50% personality labels masked (unknown)	52.1	58.5	–2.7	Sadness, happiness
Without personality nodes	50.2	56.5	–4.6	All negative emotions

be d , the number of GAT attention heads is h , and the number of graph layers is L_g . The graph attention network includes four node types: personality nodes, utterance nodes, DAR nodes, and ERC nodes. Hence, the number of nodes can be approximated as $|V_{PAU}| \approx S + 3N$. The graph attention network contains four types of edges: personality ownership edges (personality-edge), utterance temporal edges (utterance-edge), DAR edges, and ERC edges, which are used to simulate inter-task associations. Meanwhile, for temporal dependencies, we define four relations—same-personality/different-personality $\times$ past/future—to model bidirectional temporal dependencies under different personalities.

Consequently, if temporal connections are not sparsified, then in the worst case, the number of temporal edges may grow quadratically with dialogue length, i.e., $|E| \in \mathcal{O}(N^2)$. By contrast, personality ownership edges increase linearly with the number of utterances, i.e., $\mathcal{O}(N)$.

5.1.2. Time overhead of multi-graph graph attention network

For any single graph, the main overhead of a one-layer GAT comes from computing attention weights for each edge and aggregating information from neighboring nodes. If h attention heads are used, the time complexity of a single GAT layer is: $T_{\text{GAT}} = \mathcal{O}(h \cdot |E| \cdot d) + \mathcal{O}(|V| \cdot d^2)$, where the first term corresponds to edge-wise attention computation and message passing, and the second term corresponds to the node-wise linear transformation. When the graph is sparse and $|E|$ dominates, the per-layer update can be approximated as $\mathcal{O}(h \cdot |E| \cdot d)$.

In PEDR-GAT, the inference layer runs three graphs simultaneously (PAUTGAT, PDITGAT, PUDTGAT). Let their edge sets be $|E_1|$, $|E_2|$, and $|E_3|$, respectively. Then the total complexity of the GAT part in the inference layer is $T_{3\text{GAT}} = \sum_{m=1}^3 L_g \cdot (\mathcal{O}(h \cdot |E_m| \cdot d) + \mathcal{O}(|V_m| \cdot d^2))$. In the worst case, $|E_m| = \mathcal{O}(N^2)$, and thus $T_{3\text{GAT}}$ grows quadratically with the dialogue length. As the dialogue becomes longer, this term can become a bottleneck. Conversely, if temporal dependencies are sparsified, $|E_m|$ can be reduced to approximately $\mathcal{O}(Nk)$, making the graph-update complexity close to linear growth in N .

5.1.3. Time overhead of multi-head attention and the decoder

After the graph GAT updates, the model further applies multi-head attention for feature extraction and uses a GRU for global fusion. If we view this as self-attention over a sequence of graph nodes, the typical overhead of multi-

head attention is: $T_{\text{MHA}} = \mathcal{O}(h \cdot |V|^2 \cdot d)$, whose quadratic term can significantly increase inference cost in long dialogues. GRU fusion is generally linear with respect to sequence length $T_{\text{GRU}} = \mathcal{O}(|V| \cdot d^2)$.

In the classification layer, the decoder module contains multi-head self-attention and co-attention. If the decoder sequence length is aligned with the number of utterances and there are L_d stacked decoding blocks, the decoding overhead can be approximated as $T_{\text{Dec}} = \mathcal{O}(L_d \cdot h \cdot N^2 \cdot d)$.

Overall scalability with respect to N is constrained, in the worst case, by the quadratic terms from both GAT edge density and attention/decoder computations.

5.2. Model limitations and practical application

5.2.1. Limitations of personality annotation and personality modeling

Despite the effectiveness of the proposed model, several limitations arise from personality annotation, representation assumptions, and data characteristics in practical settings:

- (i) Dependence on personality annotations and noise propagation: One core assumption of the model is that personality is a stable factor participating in the construction of emotion–behavior sequences, and the Big Five traits are used as explicit inputs and graph-structured nodes. However, in real scenarios, high-quality Big Five labels are often difficult to obtain.
- (ii) Annotation sparsity and data imbalance: Case analysis suggests that inconsistencies may stem from limited Big Five annotations and imbalanced data. This implies that when personality coverage is insufficient, or when certain personality combinations appear frequently in training, the model may learn biased transfer patterns, harming generalization.
- (iii) Static assumption of personality representation: Big Five traits are commonly treated as relatively stable personal characteristics, but in real dialogues, emotional states, roles, tasks/goals, and stress can all influence behavioral expression. The interaction patterns across contexts may differ, and the model may fail to capture such contextual variations, limiting its robustness across domains and scenarios.

5.2.2. Interpretation limits of correlation patterns

Beyond performance gains, the learned personality–emotion relationships require careful interpretation due to inherent correlation and representational biases.

- (i) Inter-trait correlations and profile-based learning: The Big Five personality traits are not orthogonal psychological constructs; rather, they exhibit well-

documented intercorrelations (e.g., the typically negative association between neuroticism and extraversion, or the positive link between openness and agreeableness in certain contexts). Consequently, when PEDR-GAT incorporates personality nodes into the graph structure, the model may not be learning the distinct causal effects of individual traits (e.g., “high neuroticism independently causes sadness expression”), but rather latent trait profiles or combinatorial patterns (e.g., “the high-neuroticism/low-extraversion profile correlates with withdrawal behaviors”). For instance, the improved discrimination between anger and disgust observed in our ablation study may stem from the model’s capture of a trait cluster (low agreeableness combined with high conscientiousness) rather than the isolated effect of agreeableness. Thus, the attention weights associated with individual personality dimensions should be interpreted as indicators of profile salience rather than isolated causal factors.

- (ii) **Shortcut learning via linguistic style:** A more subtle limitation concerns shortcut learning—the risk that the model leverages surface-level linguistic markers correlated with personality traits as proxies for genuine emotional mechanisms, rather than modeling the deep psychological influence of personality on affective states. For example, speakers high in openness to experience often employ more complex vocabulary and diverse syntactic structures; the model might exploit these stylistic features to predict emotions such as “surprise” or “curiosity” without truly understanding the affective content. Similarly, the predicted personality embeddings used for DailyDialog and Mastodon effectively introduce an additional latent linguistic layer: the model may be capturing “language style typical of high-neuroticism individuals” rather than “the emotional consequences of neurotic disposition.” This distinction is critical because, while stylistic shortcuts may improve classification accuracy, they compromise the model’s generalizability to cross-cultural or cross-domain settings where personality-language associations differ.
- (iii) **Causal-inference limitation of personality modeling:** This study leverages static self-reported personality scores as prior knowledge. The model learns statistical associations between personality traits and specific patterns of emotional and behavioral expression, rather than strict causal effects. To disentangle genuine personality-driven emotional mechanisms from spurious correlations, future work should move beyond associative learning toward causal intervention

and counterfactual analysis.

5.2.3. Challenges introduced by the multi-graph structure

While the multi-graph architecture enhances expressive capability, it also introduces challenges related to computational efficiency, deployment consistency, and interpretability.

- (i) **Structural complexity and long-dialogue scalability:** The inference layer uses three graphs and additionally introduces multi-head attention and GRU fusion. As dialogue length grows or the number of participants increases, graph density and quadratic attention terms increase markedly, resulting in higher training costs and increased inference latency.
- (ii) **Offline global modeling versus online inference conflict:** This paper explicitly models bidirectional past–future temporal relations. This is beneficial for offline evaluation with full dialogue context. We acknowledge that, in real-time dialogue systems, online deployment typically has access only to historical and current information. Using a bidirectional structure directly may prevent the use of future information during deployment or require segment-wise recomputation, thereby incurring additional latency.
- (iii) **Interpretability and debugging difficulty:** We acknowledge that multi-graph, multi-edge-type modeling with attention-based fusion improves expressive power, but it also increases the difficulty of fault localization and interpretability. When performance drops, it becomes hard to quickly determine whether the issue originates from noisy personality inputs, graph construction strategies, or attention-based decoding/fusion.

5.2.4. Challenges for applying to dialogue systems

Applying personality-aware dialogue models in real-world systems introduces additional challenges related to privacy, data availability, and fairness considerations.

- (i) **Personality acquisition and privacy compliance:** If a dialogue system uses Big Five personality information, it must address privacy and compliance risks in personality inference and utilization. Personality belongs to high-sensitivity user profiling; its collection and use should comply with user consent requirements, data security, and storage requirements. Even if personality is inferred only from text, inferring such attributes without explicit consent may still raise ethical concerns.
- (ii) **Cold start and domain shift:** When new users or new speakers lack personality data, the model needs

a fallback strategy; otherwise, missing personality may invalidate the dependency anchor nodes in the graph structure. The dialogue style and distribution of training data can differ substantially from real-world business scenarios, potentially biasing the learned “personality–emotion–behavior” transfer patterns.

- (iii) Bias and fairness risks: If personality annotations are obtained manually or via automatic inference, labeling policies, cultural differences, and distribution shifts may introduce systematic biases. Treating personality as a causal factor and injecting it into the model may amplify such biases in downstream behavior prediction, affecting fairness across different user groups.
- (iv) Lack of personality annotation: Our model achieves its best performance on CPED, which validates the potential of the personality-aware architecture. However, in real-world dialogue systems, users’ personality labels are typically unknown. To address this gap, we explore two practical pathways: (a) using a neutral vector as a default personality input—under this setting, the model can still function effectively through its strong contextual modeling capacity, but the benefits of personalization are inherently limited; and (b) incorporating a lightweight personality prediction module as a front-end to provide approximate personality cues. A critical open direction is self-supervised or interactive personality inference for label-free cold-start profiling.

5.3. Limitations of personality annotations in chinese psychological emotion dialogue

The CPED dataset provides valuable Chinese dialogue data with personality information; however, models built upon it should treat the properties and limitations of its personality annotations with caution when generalizing to real-world scenarios.

- (i) Annotation source and reliability. In CPED, personality labels are obtained from self-report questionnaires. While self-report is a standard practice in personality research, its ecological validity in conversational settings remains to be verified. Whether the personality traits reported on questionnaires fully correspond to those manifested in specific dialogue situations remains an open question.
- (ii) Challenges to the static-trait assumption in Chinese contexts. Chinese communication emphasizes face and situational appropriateness. Speakers may substantially adjust their verbal behavior depending on the interlocutor’s identity and social context, which introduces tension with the common modeling

assumption that Big Five traits are fixed and invariant throughout a conversation.

- (iii) Potential cultural bias of the Big Five framework. The Big Five model originates from Western psychology. Although it has been widely adopted worldwide, subtle differences may exist in its factor structure and in how traits relate to behavior in Chinese contexts. Treating it as a universal standard directly may introduce cultural bias.

6. Conclusion

This paper proposed PEDR-GAT, a personality-aware graph attention framework for joint emotion recognition in conversation and dialogue act recognition. The model introduces three complementary graph structures—PAUTGAT, PDITGAT, and PUDTGAT—to capture personality-conditioned semantic coherence, cross-task temporal interactions, and integrated utterance–task dependencies. Experiments on Mastodon, DailyDialog, and CPED demonstrate consistent improvements over strong baselines, with F1 gains of up to 5.3% for emotion recognition and 2.3% for act recognition. Parameter-efficiency analysis confirms that performance gains stem from architectural design rather than capacity increase, and error-pattern triangulation reveals functional specialization of each graph component. While personality information improves joint-task modeling, the effects remain correlational rather than causal, and generalization to unlabeled datasets relies on linguistic style proxies. Future work should explore dynamic personality adaptation, causal inference frameworks, and lightweight architectures for real-time deployment.

Acknowledgments

None.

Funding

This work was supported by the National Natural Science Foundation of China (No. 62166029); the Research Infrastructure and Platforms in the Inner Mongolia Autonomous Region (No. 2025KYPT0014); the Inner Mongolia Autonomous Region Higher Education Innovation Team Development Plan (NO. NMGIRT2506); the Science and Technology Plan Project of Inner Mongolia Autonomous Region (No. 2025SYFHH1239); the Philosophy and Social Sciences Planning Project of Xuchang City (No.XCSK20260715).

Conflict of interest

The authors declare no competing interests.

Author contributions

Conceptualization: Zhiqiang Ma, Yutong Zhou

Data curation: Jia Liu

Formal analysis: Yutong Zhou, Kai Lv

Funding acquisition: Zhiqiang Ma

Investigation: Yutong Zhou, Kai Lv, Biqi Xu

Methodology: Zhiqiang Ma, Yutong Zhou

Project administration: Zhiqiang Ma

Resources: Zhiqiang Ma

Supervision: Zhiqiang Ma

Validation: Yutong Zhou, Kai Lv

Visualization: Yutong Zhou, Jia Liu

Writing–original draft: Yutong Zhou

Writing–review & editing: Zhiqiang Ma, Yutong Zhou, Biqi Xu

Ethics approval and consent to participate

Not applicable. This study is a secondary analysis of publicly available benchmark datasets (CPED, DailyDialog, and Mastodon). Specifically, CPED was constructed from dialogues in 40 publicly released Chinese TV dramas and annotated by the dataset creators; no human participants were recruited for this study, and no new data were collected. The authors did not conduct human subject testing, nor did we access, use, or process any identifiable private information or real-world user data. Furthermore, no testing was performed on vulnerable or special populations (e.g., minors, individuals with mental health conditions), and no assessment of potential human harm or individualized psychological impact was conducted. All datasets employed are anonymized and publicly accessible for research purposes. Therefore, ethics committee/IRB approval and informed consent were not required.

Consent for publication

Not applicable. This study did not involve human participants, and no identifiable personal data or individual images were included in this manuscript.

Availability of data

The datasets analyzed in this study are existing benchmark datasets: Mastodon, DailyDialog, and CPED, as described in the manuscript. These datasets are available from the original sources cited in the corresponding references. No new human-subject data were collected by the authors.

Further disclosure

This manuscript has been deposited as a preprint on Research Square (doi: 10.21203/rs.3.rs-3871693/v1).

References

1. Cerisara C, Jafaritazehjani S, Oluokun A, Le HT. Multi-task dialog act and sentiment recognition on Mastodon. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018;745–754.
2. Kim M, Kim H. Integrated neural network model for identifying speech acts, predicates, and sentiments of dialogue utterances. *Pattern Recognition Letters*. 2017;101:1–5.
doi: 10.1016/j.patrec.2017.11.009
3. Qin L, Che W, Li Y, Ni M, Liu T. DCR-NET: A Deep Co-Interactive Relation Network for Joint Dialog Act Recognition and Sentiment Classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(05):8665–8672.
doi: 10.1609/aaai.v34i05.6391
4. Poria S, Majumder N, Mihalcea R, Hovy EH. Emotion recognition in conversation: research challenges, datasets, and recent advances. *IEEE Access*. 2019;7:100943–100953.
5. Xing B, Tsang I. DARER: Dual-task Temporal Relational Recurrent Reasoning Network for joint Dialog Sentiment Classification and ACT recognition. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022. 2022;3611–3621.
doi: 10.18653/v1/2022.findings-acl.286
6. Zhang Y, Zhao G, Shu Y, et al. CPED: a Chinese Positive Emotion Database for emotion Elicitation and analysis. *IEEE Trans Affect Comput*. 2021;14(2):1417–1430.
doi: 10.1109/taffc.2021.3088523
7. Xu Y, Yao E, Liu C, Liu Q, Xu M. A novel ensemble model with two-stage learning for joint dialog act recognition and sentiment classification. *Pattern Recognition Letters*. 2022;165:77–83.
doi: 10.1016/j.patrec.2022.11.032
8. Li S, Yan H, Qiu X. Contrast and generation make BART a good dialogue emotion recognizer. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2022;36(10):11002–11010.
doi: 10.1609/aaai.v36i10.21348
9. Ayetiran EF. Attention-based aspect sentiment classification using enhanced learning through cnn-Bilstm networks. *Knowl-Based Syst*. 2022;252:109409.
doi: 10.1016/j.knosys.2022.109409
10. Majumder N, Poria S, Hazarika D, Mihalcea R, Gelbukh A, Cambria E. DialogueRNN: an attentive RNN for emotion detection in conversations. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019;33 (1):6818–6825.

11. Hu D, Wei L, Huai X. DialogueCRN: Contextual Reasoning Networks for Emotion Recognition in Conversations. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. 2021;1:7042–7052.
12. Kumar H, Agarwal A, Dasgupta R, Joshi S. Dialogue act sequence labeling using hierarchical encoder with CRF. In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18)*. AAAI Press. 2018;421:3440–3447.
13. Raheja V, Tetreault J. Dialogue Act Classification with Context-Aware Self-Attention. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2019;1 (Long and Short Papers):3727–3733.
14. Duran N, Battle S, Smith J. Sentence encoding for Dialogue Act classification. *Nat Lang Eng*. 2021;29(3):794–823.
doi: 10.1017/s1351324921000310
15. Malhotra G, Waheed A, Srivastava A, Akhtar MS, Chakraborty T. Speaker and time-aware joint contextual learning for dialogue-act classification in counselling conversations. In: *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 2022;735–745.
doi: 10.1145/3488560.3498509
16. Zheng L, Li F, Chai Y, Teng C, Ji D. A bi-directional multi-hop inference model for joint dialog sentiment classification and ACT recognition. In: Liu, F, Duan, N., Xu, Q., Hong, Y. (eds) *Proceedings of the Natural Language Processing and Chinese Computing. NLPCC 2023. Lecture Notes in Computer Science*. 2023;14302.
doi: 10.1007/978-3-031-44693-1_19.
17. Li J, Fei H, Ji D. Modeling local contexts for joint dialogue act recognition and sentiment classification with bi-channel dynamic convolutions. In: *Proceedings of the International Conference on Computational Linguistics*. 2020;616–626.
18. Ghosal D, Majumder N, Poria S, Chhaya N, Gelbukh A. DialogueGCN: a graph convolutional neural network for emotion recognition in conversation. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2019;154–164.
19. Zhang D, Wu L, Sun C, Li S, Zhu Q, Zhou G. Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. 2019;5415–5421.
20. Zhao W, Zhao Y, Qin B. MuCDN: mutual conversational detachment network for emotion recognition in multi-party conversations. In: *Proceedings of the International Conference on Computational Linguistics*. 2022;7020–7030.
21. Zhang D, Chen F, Chen X. DualGATs: dual graph attention networks for emotion recognition in conversations. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2023;1:7395–7408.
22. Qin L, Li Z, Che W, Ni M, Liu T. 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*. 2021;35(15):13709–13717.
23. Li S, Chen X. Multiple Information-Aware Recurrent Reasoning Network for Joint Dialogue Act recognition and Sentiment Classification. *Information*. 2023;14(11):593.
doi: 10.3390/info14110593
24. Li Y, Su H, Shen X, Li W, Cao Z, Niu S. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing*. 2017;1:986–995.
25. Schegloff EA. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge, UK: Cambridge University Press. 2007;1.
doi: 10.1017/cbo9780511791208
26. Pennebaker JW, King LA. Linguistic styles: language use as an individual difference. *J Pers Soc Psychol*. 1999;77(6):1296–1312.
doi: 10.1037/0022-3514.77.6.1296
27. Tett RP, Guterman HA. Situation trait relevance, trait expression, and cross-situational consistency: testing a principle of trait activation. *J Res Pers*. 2000;34(4):397–423.
doi: 10.1006/jrpe.2000.2292.
28. Searle JR. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press; 1969.
29. Hatfield E, Cacioppo JT, Rapson RL. Emotional contagion. *Curr Dir Psychol Sci*. 1993;2(3):96–100.
doi: 10.1111/1467-8721.ep10770953.
30. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. In: *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*. 2017.
31. Kazameini A, Fatehi S, Mehta Y, Eetemadi S, Cambria E. Personality trait detection using bagged SVM over BERT word embedding ensembles. *arXiv* 2020. <https://arxiv.org/abs/2010.01309>

APPENDIX

A1. Detailed mathematical derivations for personality-enhanced dialogue representation graph attention network**A1.1. Feature extraction initialization layer**

$$d_i = \text{soft max}(\text{Fc}(\text{Matt}(\text{H}_{\text{pe}}, d_i^t, d_i^t))). \quad (5)$$

$$e_i = \text{soft max}(\text{Fc}(\text{Matt}(\text{H}_{\text{pe}}, e_i^t, e_i^t))). \quad (A1)$$

where d_i^t and e_i^t represent local feature representations of act labels $D_{\text{part}} = D^t = \{d_i^t\}_{i=1}^N$ and emotional labels $E_{\text{part}} = E^t = \{e_i^t\}_{i=1}^N$, respectively.

where,

$$d_i^t = \sum_{k=0}^{|C_D|-1} p_{D,i}^{t-1}[c_d] \cdot v_D^{c_d}. \quad (A2)$$

$$e_i^t = \sum_{k=0}^{|C_E|-1} p_{E,i}^{t-1}[c_e] \cdot v_E^{c_e} \quad (A3)$$

where, $v_D^{c_d}$ and $v_E^{c_e}$ represent embedded labels for act category c_d and emotional category c_e , while $p_{D,i}^t$ and $p_{E,i}^t$ denote the initial label distributions for act and emotion. The calculations are as follows.

$$P_D^0 = \{P_{D,i}^0\}_{i=1}^N, P_{D,i}^0 = \text{softmax}(W_d H_{\text{pe}}^0 + b_d) = [p_{D,i}^0[0], \dots, p_{D,i}^0(|C_D|-1)] \quad (A4)$$

$$P_E^0 = \{P_{E,i}^0\}_{i=1}^N, P_{E,i}^0 = \text{softmax}(W_e H_{\text{pe}}^0 + b_e) = [p_{E,i}^0[0], \dots, p_{E,i}^0(|C_E|-1)] \quad (A5)$$

where, W_* and b_* represent the weight matrix and bias, while C_D and C_E correspondingly denote the sets of act categories and emotion categories.

A1.2. Personality-aware utterance temporal graph

Calculation of PAUTGAT in the model involves the following steps: For any nodes x_i and x_j , place them into $G^{\text{Paut}} = (S^{\text{Paut}}, P^{\text{Paut}}, R^{\text{Paut}})$ for updating. Specifically, for a given node S^{Paut} , PAUTGAT aggregates information from its neighboring nodes as follows:

$$\beta_{[i][j]} = \text{Soft max}(\text{LeakyReLU}(a^T [Wx_i \parallel Wx_j \parallel P_i^{\text{Paut}} \parallel P_j^{\text{Paut}} \parallel R_{[i][j]}^{\text{Paut}}])) \quad (A6)$$

$$x_i^{\text{paut}} = \parallel \sigma \left(\sum_{k=1}^k \beta_{[i][j]}^k Wx_j \right) \quad (A7)$$

where, $R_{[i][j]}^{\text{Paut}} = l^{\text{Paut}}$ satisfies the discourse dependency relationship $l^{\text{Paut}} \in L^{\text{Paut}}$, $\beta_{[i][j]}$ represents the weights on the edges from the node S_i^{Paut} to its adjacent node S_j^{Paut} , W and a^T represent trainable parameters, while $\parallel$ represents the concatenation operation. N_i^{Paut} represents the neighborhood nodes of the node, $x_i^{\text{Paut}} \in \tilde{R}^{d_h}$ represents the nodes related to S_i^{Paut} after the PAUTGAT update, d_h represents the dimension of the intermediate vector. After updating all nodes, it is represented as: $X^{\text{Paut}} = \text{SPETGAT}(X, P^{\text{Paut}}, R^{\text{Paut}})$, $X^{\text{Paut}} \in \tilde{R}^{N \times d_h}$.

A1.3. Personality dual-task interaction temporal graph

Calculation of PDITGAT in the model involves the following steps: The initial local features obtained for $D^t = \{d_i^t\}_{i=1}^N$ and $E^t = \{e_i^t\}_{i=1}^N$ are placed into $G^{Pdit} = \{(D^{Pdit}, P^{Pdit,D}, R^{Pdit,D}), (E^{Pdit}, P^{Pdit,E}, R^{Pdit,E})\}$, and the nodes are updated as follows:

$$\alpha_{[i][j]}^D = \text{Soft max}(\text{LeakyReLU}(a^T[Wd_i \parallel Wd_j \parallel P_i^{Pdit,D} \parallel P_j^{Pdit,D} \parallel R_{[i][j]}^{Pdit,D}])) \quad (\text{A8})$$

$$d_i^{Pdit,t} = \parallel \sigma(\sum_{k=1}^k \sum_{j \in N_i^{Pdit,D}} \alpha_{[i][j]}^{D,k} W^k d_j^{t-1}) \quad (\text{A9})$$

$$\alpha_{[i][j]}^E = \text{Soft max}(\text{LeakyReLU}(a^T[We_i \parallel We_j \parallel P_i^{Pdit,E} \parallel P_j^{Pdit,E} \parallel R_{[i][j]}^{Pdit,E}])) \quad (\text{A10})$$

$$e_i^{Pdit,t} = \parallel \sigma(\sum_{k=1}^k \sum_{j \in N_i^{Pdit,E}} \alpha_{[i][j]}^{E,k} W^k e_j^{t-1}) \quad (\text{A11})$$

Personality dual-task interaction temporal graph interactive update is as follows:

$$d_i^{Pdit,t+1} = \parallel \sigma(\sum_{k=1}^k \sum_{j \in N_i^{Pdit,D}} \alpha_{[i][j]}^k W^k d_j^t + \sum_{j \in N_i^{Pdit,E}} \alpha_{[i][j]}^k W^k e_j^t) \quad (\text{A12})$$

$$e_i^{Pdit,t+1} = \parallel \sigma(\sum_{k=1}^k \sum_{j \in N_i^{Pdit,D}} \alpha_{[i][j]}^k W^k e_j^t + \sum_{j \in N_i^{Pdit,E}} \alpha_{[i][j]}^k W^k d_j^t) \quad (\text{A13})$$

After updating all nodes, PDITGAT is represented as: $\begin{matrix} D^{Pdit} \\ E^{Pdit} \end{matrix} = PDITGAT\left\{\begin{matrix} (D, P^{Pdit,D}, R^{Pdit,D}) \\ (E, P^{Pdit,E}, R^{Pdit,E}) \end{matrix}\right\}$

A1.4. Personality-aware utterance and dual-task integrated temporal graph

The features after GAT and the features after the multi-head attention mechanism are globally fused using GRU, represented as follows:

$$A_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (\text{A14})$$

$$H_{multi} = \text{MAtt}(Q, K, V) = [A_1, \dots, A_n] W^O \quad (\text{A15})$$

$$\tilde{D}_{over} = GRU(D^{Puda} \oplus X^{Paut} \oplus D^{Pdit}) \quad (\text{A16})$$

$$\tilde{E}_{over} = GRU(E^{Puda} \oplus X^{Paut} \oplus E^{Pdit}) \quad (\text{A17})$$

$$D_{overall} = H_{multi} + \tilde{D}_{over}, E_{overall} = H_{multi} + \tilde{E}_{over} \quad (\text{A18})$$

A1.5. Dual-task classification layer

After multiple interactions with the feature extraction initialization layer and the personality-fused interactive reasoning layer distributed and stacked with interaction layers, the model utilizes co-attention to capture important relevant information. This process produces the final output of the interaction layer. Subsequently, separate decoders are employed for predicting conversational act and emotion, represented as follows:

$$P(y^E | E_{part}^l, E_{overall}^l) = \prod_{t=1}^N p(y_t^E | y_1^E, y_2^E, \dots, y_{t-1}^E; E_{part}^l, E_{overall}^l) \quad (A19)$$

$$P(y^D | D_{part}^l, D_{overall}^l) = \prod_{t=1}^N p(y_t^D | y_1^D, y_2^D, \dots, y_{t-1}^D; D_{part}^l, D_{overall}^l) \quad (A20)$$

where $E_{part}^l = (e_{part1}^l, e_{part2}^l, \dots, e_{partN}^l)$ and $D_{part}^l = (d_{part1}^l, d_{part2}^l, \dots, d_{partN}^l)$ are sequences generated by the personality-guided feature extraction initialization layer. $E_{overall}^l = (e_{overall1}^l, e_{overall2}^l, \dots, e_{overallN}^l)$ and $D_{overall}^l = (d_{overall1}^l, d_{overall2}^l, \dots, d_{overallN}^l)$ are sequences generated by the personality-fused interactive reasoning layer.

The decoder consists of l identical blocks, each containing a multi-head self-attention layer, a co-attention layer, and a feedforward layer. In the emotion decoding process, the outputs of the first sub-layer $M_{E,part}^l$ and $M_{E,overall}^l$, the second sub-layer C_E^l , and the third sub-layer E^l of the l -th decoding block are calculated as follows:

$$M_{E,part}^l = LN(SATT(E_{part}^{l-1}) + E_{part}^{l-1}) \quad (A21)$$

$$M_{E,overall}^l = LN(SATT(E_{overall}^{l-1}) + E_{overall}^{l-1}) \quad (A22)$$

$$C_E^l = LN(Co-Att(M_{E,part}^{l-1}, M_{E,overall}^{l-1}) + M_{E,part}^{l-1} + M_{E,overall}^{l-1}) \quad (A23)$$

$$Co-Att() = M_{E,overall}^{l-1} + Soft \max(M_{E,overall}^{l-1} ((M_{E,part}^{l-1})^N) M_{E,part}^{l-1}) \quad (A24)$$

$$E^l = LN(FFN(C_E^l) + C_E^l) \quad (A25)$$

where LN represents the normalization layer; $SATT$ represents a multi-head self-attention mechanism, $Co-Att$ represents the co-attention layer, FFN represents a feedforward neural network. The decoding for conversational act prediction follows the same process.

Finally, utilizing the output $D^l = (d_1^l, d_2^l, \dots, d_N^l)$ and $E^l = (e_1^l, e_2^l, \dots, e_N^l)$ from the last layer of the decoder, the generation of features at time t can be formalized as:

$$y_t^E = soft \max(W^E e_t^l + I_t^E) \quad (A26)$$

$$y_t^D = soft \max(W^D d_t^l + I_t^D) \quad (A27)$$

where I_t^E and I_t^D are mask vectors at the decoding step t used to prevent the decoder from attending to repeated features, W^E and W^D are transformation matrices for conversation emotion and act, and l is the number of stacked relation layers in our framework.

A2. Supplemental sequence-level evaluation with baseline language understanding evaluation–emotion prediction

The baseline language understanding evaluation–emotion prediction (BLUE-EP) is an automatic evaluation metric used to assess the degree of fit between the perceived and original sequences. The BLUE-EP is used to analyze the overlap accuracy between candidate and reference sequences, following the BLUE metric proposed by International Business Machines Corporation in 2002. BLUE-EP calculation is as follows.

$$EP_n = \frac{\sum_i \sum_k \min(h_k(ep_i), \max_{j \in m} h_k(e_{ij}))}{\sum_i \sum_k h_k(ep_i)} \quad (A28)$$

$$b_e = \begin{cases} 1 & \text{if } l_{ep} > l_e \\ e^{1 - \frac{l_e}{l_{ep}}} & \text{if } l_{ep} \leq l_e \end{cases} \quad (A29)$$

$$BLUE-EP = b_e \exp\left(\sum_{n=1}^N s_n \log EP_n\right) \quad (A30)$$

In this equation, ep_i represents the candidate sequence corresponding to the reference sequence $E_i = \{e_{i1}, e_{i2}, \dots, e_{ij}\}$. The reference n -gram is defined as a set of sequences of length l , where S_k represents k groups of n -grams. $h_k(ep_i)$ and $h_k(e_{ij})$ respectively denote the occurrences of S_k in the candidate and reference sequences. l_{ep} and l_e represent the effective lengths of the candidate sequence ep_i and reference sequence e_{ij} . Typically, s_n is assigned a constant value of $\frac{1}{n}$.

To analyze the degree of overlap between the sequences obtained after emotion and behavior perception and the reference sequences, comparative experiments were conducted using the BLUE-EP metric on the CPED and DailyDialog datasets. The experimental results are shown in Table A1. In the CPED dataset, the fitting BLUE-EP of the emotion sequence improved by 1.8% compared to the best baseline (DARER), while the fitting BLUE-EP of the behavior sequence decreased by 1.1% compared to the best baseline (DARER). In the DailyDialog dataset, the fitting BLUE-EP of the emotion sequence increased by 2.6% compared to the best baseline (DARER), and the fitting BLUE-EP of the behavior sequence increased by 1.2%. The performance improvement can be attributed to two factors: PEDR-GAT can learn dependencies between emotion and behavior sequences within the same personality traits across different speakers, and it can capture temporal dependencies between emotion and behavior sequences to mutually promote each other.

A3. Error severity analysis (practical significance analysis)

To supplement macro-averaged F1 metrics, we define a severity taxonomy of errors based on psychological impact and application-level consequences, as shown in Table A2. Furthermore, we compared PEDR-GAT against the baseline (DARER) on the CPED test set, with results shown in Table A3.

Critical error reduction focuses on high-neuroticism speakers (84.6% of prevented distress→happiness errors), where PDITGAT's cross-task coupling detects inconsistencies between positive emotion labels and help-seeking action patterns (e.g., "I am fine" + "Can you stay with me?"). This validates that personality-aware modeling prevents pragmatically dangerous misclassifications that semantic-only models miss.

Table A1. BLUE-EP scores of PEDR-GAT on CPED and DailyDialog datasets

Model	CPED		DailyDialog	
	ERC	DAR	ERC	DAR
JointDAS	0.451	0.438	0.470	0.684
IIIM	0.475	0.493	0.482	0.702
DCR-Net	0.461	0.502	0.505	0.720
BCDCN	0.453	0.498	0.512	0.723
Co-GAT	0.506	0.532	0.527	0.744
DARER	0.554	0.576	0.559	0.774
PEDR-GAT	0.561	0.582	0.547	0.756

Abbreviations: BCDCN: Bi-channel dialogue context network; BLUE-EP: Baseline language understanding evaluation–emotion prediction; CPED: Chinese psychological emotion dialogue; Co-GAT: Conversational graph attention network; DAR: Dialogue act recognition; DARER: Dialogue act recognition and emotion recognition; DCR-Net: Deep collaborative relationship network; ERC: Emotion recognition in conversations; IIIM: Interactive information integration model; JointDAS: Joint dialogue act and sentiment recognition; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition.

Table A2. Severity taxonomy and operationalization

Severity level	Definition	Example errors	Application impact
Critical	Valence inversion or safety-relevant omission	Distress→happiness	User harm; system failure
High	Arousal-category confusion within the same valence	Anger→disgust; Fear→sadness	Misaligned response
Moderate	Neutral-emotion misclassification	Neutral→sadness; Happy→neutral	Reduced engagement
Low	Fine-grained category error within the same arousal/valence	Relaxed→content	Minimal impact

Table A3. Critical error reduction (CPED test set)

Error type	DARER (%)	PEDR-GAT (%)	Reduction
Distress → happiness	4.2	2.5	40
Suicidal cue → neutral	1.8	1.0	44
Anger → neutral	8.6	6.4	25.5
Fear → neutral	6.3	4.8	23.8
Weighted critical error rate	5.2	3.6	30.7

Abbreviations: CPED: Chinese psychological emotion dialogue; DARER: Dialogue act recognition and emotion recognition; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition.

A4. Statistical testing (paired permutation test)

Statistical testing protocol. We assessed significance using paired permutation tests (10,000 permutations, $\alpha = 0.05$). “Pairs” are defined as matched predictions on identical test-set utterances from two models. We tested macro-F1 for ERC and DAR separately, conducting 18 comparisons (3 datasets $\times$ 2 tasks $\times$ 3 primary baselines) with Bonferroni correction ($\alpha_{\text{adj}} = 0.05/18 \approx 0.0028$). 95% bootstrap confidence intervals (10,000 resamples, percentile method) for key comparisons are reported in [Table A4](#).

Table A4. Bootstrap confidence intervals for key comparisons

Comparison	Dataset–task	F1 difference	95% CI	<i>p</i> -value (adjusted)
PEDR-GAT versus DARER	CPED–ERC	+5.3%	[+3.2, +7.4]	0.0012*
PEDR-GAT versus DARER	CPED–DAR	+2.3%	[+0.8, +3.9]	0.0087*
PEDR-GAT versus DARER	DailyDialog–ERC	+3.3%	[+1.5, +5.1]	0.0045*
PEDR-GAT versus DARER	DailyDialog–DAR	+1.3%	[+0.2, +2.5]	0.0312
PEDR-GAT-C versus DARER	CPED–ERC	+1.2%	[−0.5, +2.9]	0.1823

Note: * denotes statistical significance at $p < 0.05$.

Abbreviations: CPED: Chinese psychological emotion dialogue; DAR: Dialogue act recognition; DARER: Dialogue act recognition and emotion recognition; ERC: Emotion recognition in conversation; PEDR-GAT: Personalized model integrating personality for joint dialogue emotion classification and act recognition; PEDR-GAT-C: Personalized model integrating personality for joint dialogue emotion classification and act recognition with context modeling.