International Journal of Innovative Computing, Information and Control
Volume **21**, Number **6**, December **2025**

ICIC International ©2025 ISSN 1349-4198
pp. **1583–1597**

JOINT MODEL FOR MULTI-INTENT SPOKEN LANGUAGE UNDERSTANDING BASED ON BIDIRECTIONAL GRAPH ATTENTION NETWORK AND ENHANCED WITH LARGE LANGUAGE MODELS

Shuxin Chen[1], Xu Li[2,*], Jiaqi Wang[1] and Yu Zhang[1]

[1]School of Information Science and Engineering
[2]Innovation and Entrepreneurship Education Center
Dalian Polytechnic University
No. 1, Qinggong Yuan, Ganjingzi District, Dalian 116034, P. R. China
{ chenshuxin; wangjiaqi; zhangyu }@dlpu.edu.cn
*Corresponding author: lixu@dlpu.edu.cn

Received April 2025; revised July 2025

Abstract. *Spoken language understanding is a critical task in task-oriented dialogue systems, primarily composed of two subtasks: semantic slot filling and intent recognition, which are highly correlated and typically trained jointly. Since multiple intent information in utterances can guide semantic slot filling, and semantic slot information can also assist in better intent recognition, the model employs a graph attention network to establish bidirectional interactions between intents and semantic slots. Additionally, the large language model ChatGPT is introduced to effectively capture complex semantic associations and long-range dependencies in dialogues, enhancing the model's contextual awareness. The experimental results on the MixATIS and MixSNIPS datasets show that our method achieves semantic accuracy improvements of 1.0% and 0.8%, respectively, compared to previous models.*

Keywords: Intent recognition, Slot filling, Graph attention network, Large language model

1. **Introduction.** Spoken Language Understanding (SLU) [1] is a critical component of task-oriented dialogue systems. It typically includes two classical subtasks: intent detection and slot filling. Intent detection focuses on predicting the user's intent, which can be regarded as a text classification problem. Slot filling focuses on extracting entity information from the text, predicting entity type labels for each word in the utterance, and this task is generally considered a sequence labeling problem [2]. Early approaches treated intent detection and slot filling as separate tasks without considering their interdependencies. However, in real-world scenarios, users often express multiple intents within a single utterance. Taking the example in Figure 1, the utterance indicates two intents: "PlayMusic" and "GetWeather". Therefore, researchers have begun to adopt multi-task learning principles to construct joint frameworks for multi-intent detection and slot filling. Among these, Gangadharaiah [3] first proposed a joint framework for multi-intent detection and slot filling under a multi-task framework, treating multi-intent detection as a multi-label classification task. They introduced sentence-level and token-level joint multi-intent detection and employed a gating mechanism to model dependencies between intents and semantic slots. However, the gating mechanism failed to adequately summarize and capture intent information. To address this, Qin et al. [4] proposed an adaptive graph-interaction framework (AGIF) for joint multi-intent detection and semantic slot filling, utilizing a

DOI: 10.24507/ijicic.21.06.1583

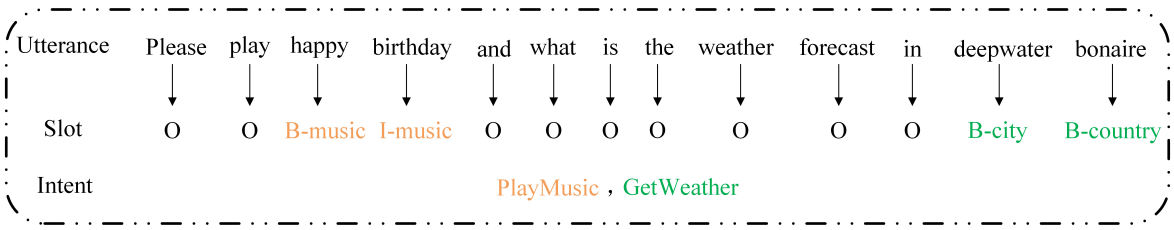

FIGURE 1. An example of utterance with multiple intents and slots

graph neural network architecture to establish strong correlations between intents and semantic slots. Building on this, in 2021, Qin et al. [5] further employed graph neural networks to construct local semantic slot interaction graphs and global intent-semantic slot interaction graphs to model the connections between the two tasks.

Although existing joint models for multi-intent Spoken Language Understanding have achieved significant progress, two major issues remain.

1) Neglect of guidance from slots to intents. While researchers acknowledge the dependency of slot labels on intents [3], existing models primarily utilize predicted intent information to guide slot filling but fail to fully exploit the guidance value of slot information for multi-intent detection. Observations reveal that multi-intent detection and slot filling can mutually instruct each other. In previous joint models, multi-intent detection could only receive guidance from shared semantic features of slots. The lack of bidirectional guidance from slots to intents restricts the improvement of model accuracy. To address this, this study constructs a bidirectional interaction architecture based on a graph attention network [6], leveraging GAT's capability to explicitly distinguish intent and slot nodes through heterogeneous graph construction [7], enable localized attention focusing on task-specific dependencies [8], and support dynamic bidirectional message passing [9]. This allows the model to integrate multi-intent information from utterances to guide relevant slot filling while leveraging slot information to enhance intent recognition, addressing the underutilization of task-related associations in existing joint models and strengthening their interactivity. 2) Traditional joint modeling approaches for multi-intent recognition and slot filling often exhibit a disconnect between intents and slots when dealing with complex contexts and long textual expressions, resulting in incoherent semantic expressions and loose logical relationships. Additionally, reliance on predefined rules and limited domain knowledge hinders their ability to capture fine-grained semantic distinctions and cross-task dependencies, making them inadequate for dynamic linguistic environments and nested complex semantics. To resolve these limitations, our method introduces ChatGPT for preprocessing raw text. This approach fully leverages large language models' deep semantic understanding and open-domain knowledge to comprehensively optimize text semantics at the preprocessing stage, thereby providing more precise and coherent inputs for subsequent joint modeling while significantly enhancing the system's generalization capability and overall performance in complex scenarios [10,11].

Experiments conducted on two public multi-intent utterance datasets demonstrate that our method achieves markedly superior overall utterance accuracy compared to previous state-of-the-art models.

2. **Related Work.** In the field of SLU, research on joint modeling of intent detection and slot filling has undergone continuous technological evolution. In early work, Liu and Lane [12] pioneered the integration of attention mechanisms into recurrent neural network (RNN) architectures to explore the enhancement of explicit alignment information for slot

filling tasks. Subsequently, Goo et al. [13] proposed the slot-gated mechanism, establishing a new paradigm for guiding slot prediction with intent information. Li et al. [14] and E et al. [15] further improved joint model performance by refining feature interaction methods. Addressing the pervasive challenge of multi-intent scenarios in real-world dialogues, Kim et al. [16] first developed a multi-intent SLU system, while Gangadharaiah [3] innovatively implemented joint multi-intent detection and slot prediction through slot-gate mechanism. Their study on Amazon internal data revealed that 52% of dialogues involve multiple intents, underscoring the research significance of this direction.

During technological advancements, Wang et al. [17] broke the unidirectional guidance paradigm by constructing a bidirectional attention interaction module to enable mutual information transfer between intents and slots. Zhang et al. [18] addressed the long-range dependency limitations of RNN by introducing graph neural network (GNN) to joint modeling tasks for the first time. The Qin's team [4] proposed an adaptive graph interaction framework based on GAT in 2020 to achieve fine-grained multi-intent feature fusion, and further developed a non-autoregressive global-local graph network in 2021 [5], significantly enhancing model efficiency through parallel decoding. Recent studies demonstrate diversified breakthroughs: Chen et al. [19] and Chen et al. [20] reformulated multi-intent detection as a weakly supervised task and designed a self-distillation mechanism for iterative optimization; Song et al. [21] constructed a global graph model to capture intent-slot co-occurrence relationships; Cheng et al. [22] proposed the SSRAN model, employing a span-sensitive attention network to mitigate error propagation. Progress has also been made in dialect processing, with Abboud and Oz [23] generating synthetic training data for German and Arabic dialects via masked language modeling, while Joshi et al. [24] systematically outlining a framework for dialect natural language processing (NLP) research.

This work builds upon and extends the aforementioned technical framework by constructing a dual-interaction graph structure and applying a graph attention network to achieving deep mutual guidance between intent and slot tasks. Additionally, it incorporates the large language model ChatGPT to transcend the limitations of traditional methods in semantic representation. Experimental results demonstrate that this approach significantly improves the overall performance of intent recognition and slot filling.

3. **Methodology.**

3.1. **Problem definition.** Given an input utterance $U = [u_1, u_2, \ldots, u_n]$, where n denotes the utterance length, the joint multi-intent detection and slot filling task comprises two subtasks: 1) Multi-intent detection is defined as a multi-label classification task aiming to predict multiple intent labels $O_I = [o_I^1, o_I^2, \ldots, o_I^m]$ corresponding to the input utterance, where m represents the number of output intent labels; 2) Slot filling is defined as a sequence labeling task that assigns the corresponding slot label to each word u_i in the input utterance, outputting the slot label sequence $O_S = [o_s^1, o_s^2, \ldots, o_s^n]$. To enhance inter-task correlations, we propose an intent-slot bidirectional interaction model, as illustrated in Figure 2. The framework consists of four key components: 1) Utterance encoder; 2) Bidirectional interactive graph attention network for modeling intent-slot dependencies; 3) Independent decoders for intent detection and slot filling; 4) ChatGPT integration to handle complex multi-intent and contextual dependency scenarios, significantly boosting overall task performance through tight coupling with the core model. Under a joint training strategy, the model simultaneously optimizes both intent detection and slot filling tasks, thereby enhancing the holistic performance of spoken language understanding.

3.2. **Utterance encoder.** Following Qin et al. [4] and Qin et al. [5], to obtain semantic representations of input utterances, we employ a task-shared and task-specific encoder

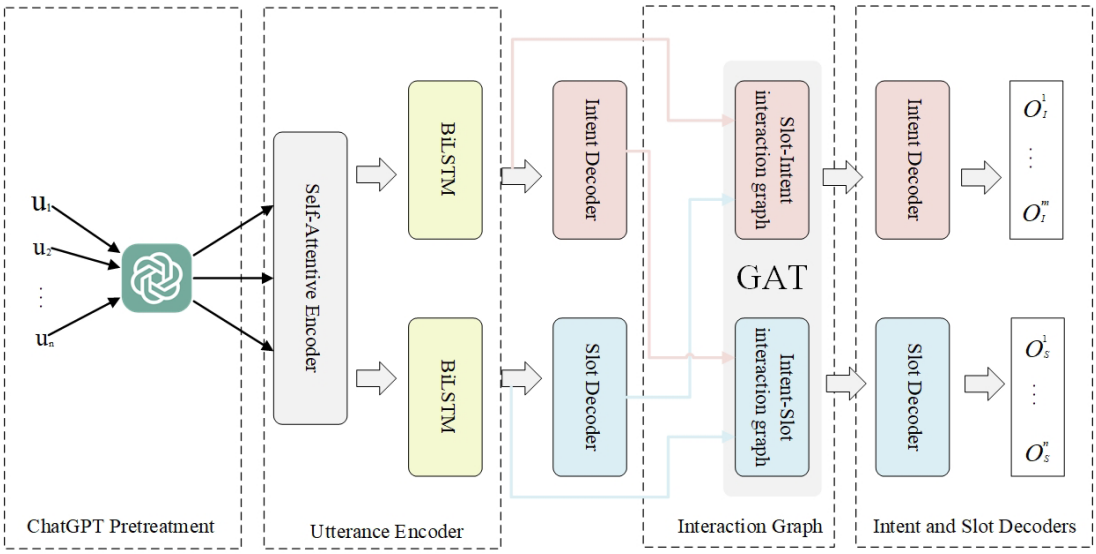

FIGURE 2. The architecture of our proposed approach

framework. First, the task-shared encoder captures global contextual information of the input utterance through a bidirectional LSTM (BiLSTM) and self-attention mechanisms. Specifically, the BiLSTM models dependencies by processing the input sequence in both forward and backward directions:

$$h_i = \text{BiLSTM}(x_i, h_{i-1}, h_{i+1}) \tag{1}$$

Generate context-sensitive hidden states $H = \{h_i\}_{i=1}^{n}$, where x_i is the word embedding vector of u_i.

Subsequently, the self-attention mechanism applies distinct linear projections to the word vector matrix of the input utterance, generating Query (Q), Key (K), and Value (V) matrices to capture global dependency relationships:

$$C = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{2}$$

Output the contextual representation C. Finally, concatenate the outputs of the BiLSTM and self-attention to obtain the task-shared encoded representation $E = H \| C$.

In the task-specific encoder, we separately model the task-shared representation E using an intent-aware BiLSTM and a slot-aware BiLSTM. For intent detection, E is input into the intent-aware BiLSTM to generate the task-specific intent representation E^{Intent}; for slot filling, the slot-aware BiLSTM outputs the task-specific slot representation E^{Slot}. By combining task-shared and task-specific approaches, the model can effectively capture contextual features of both intents and slots, strengthen the correlations between tasks, thereby enhancing the performance of joint intent detection and slot filling.

3.3. **Intent-slot graph interaction layer.** The proposed intent-slot graph interaction layer is built upon a GAT to establish a bidirectional information interaction mechanism. It dynamically associates multi-intent and slot information to enable collaborative reasoning. Specifically, the model leverages the masked self-attention layer of GAT to deeply integrate graph structural information with node features, allowing each node to adaptively learn the importance weights of its neighboring nodes and automatically capture local semantic associations. Given a graph structure with N nodes, a one-layer GAT takes the initial node feature set $Z = \{z_1, z_2, \ldots, z_n\}$ $\left(z_n \in \mathbb{R}^F\right)$ as input and generates more abstract representations $Z' = \{z'_1, z'_2, \ldots, z'_n\}$ $\left(z'_n \in \mathbb{R}^{F'}\right)$ through an attention mechanism.

The attention mechanism of GAT can be summarized as follows:

$$f(z_i, z_j) = \text{LeakyReLU}\left(a^\top \left[W_z z_i \| W_z z_j\right]\right) \tag{3}$$

$$z_i' = \sigma\left(\sum_{j \in N_i} a_{ij} W_z z_j\right) \tag{4}$$

$$a_{ij} = \frac{\exp\left(f(z_i, z_j)\right)}{\sum_{j' \in N_i} \exp\left(f(z_i, z_{j'})\right)} \tag{5}$$

where N_i is the first-order neighbors of node i (including i) in the graph, $W_z \in \mathbb{R}^{F' \times F}$ and $a \in \mathbb{R}^{2F'}$ is the trainable weight matrix, a_{ij} is the normalized attention weight and σ represents the nonlinearity activation function.

The bidirectional interaction layer between intents and semantic slots based on the graph attention network constructed in Figure 2 is specifically illustrated in Figure 3 and Figure 4. This layer achieves deep collaborative modeling between intents and slots, enabling dynamically associated predictions of multi-intent labels and semantic slot nodes through graph structures. Specifically, in Figure 3, the nodes include slot label nodes (S_1, S_2) from the slot decoder and intent semantic nodes (i_1', i_2') from the BiLSTM. Their interactions are reflected through the edge relationships in the graph. The connections in the graph include 1) Connections between slot label nodes: The bidirectional connections between S_1 and S_2 represent relationships between slot label nodes. These connections reflect mutual dependencies among slot labels and enhance contextual associations between slot labels through the graph structure, particularly in multi-slot annotation tasks, enabling the capture of dependencies between slots. 2) Connections between intent semantic nodes: The bidirectional connections between i_1' and i_2' indicate associations between intent semantic nodes. Through these connections, the model can model potential dependencies between different intent semantics and improve the accuracy of multi-intent recognition via the information propagation mechanism of the graph structure. 3) Connections between intent semantic nodes and slot label nodes: The bidirectional edges between S_1, S_2 and i_1', i_2' represent associations between intent semantic nodes and slot label nodes. Through these connections, the graph structure enables cross-task interactions between intent information and slot labels, capturing potential associations between intent semantics and slot labels. This further enhances the collaborative effects between intent detection and slot filling tasks.

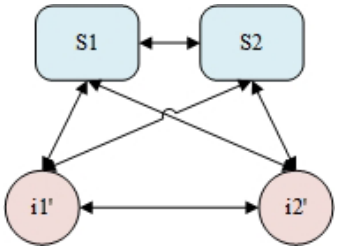

FIGURE 3. Slot-Intent interaction graph

The nodes in Figure 4 include intent label nodes (I_1, I_2) from the intent decoder and slot semantic nodes (s_1', s_2') from the BiLSTM output. Three primary edge connections exist in the graph: 1) Connections between intent label nodes: The edge between I_1 and I_2 represents relationships between intent label nodes. Their interconnections through the graph structure model potential associations between multiple intent labels within the same sentence. 2) Connections between slot semantic nodes: The edge between s_1'

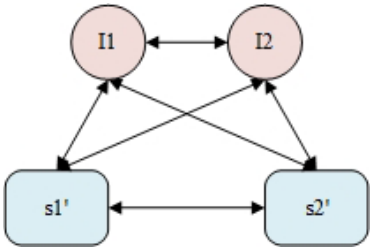

FIGURE 4. Intent-Slot interaction graph

and s_2' indicates associations between slot semantic nodes. These connections enhance semantic coherence in slot filling tasks and help capture local contextual dependencies. 3) Connections between intent label nodes and slot semantic nodes: The bidirectional edges between I_1, I_2 and s_1', s_2' represent cross-task interactions between intent label nodes and slot semantic nodes. Through the graph structure, these connections propagate intent information and slot information, further strengthening the semantic associations between intents and slots.

These connections collectively form a collaborative modeling framework for multi-intent and multi-slot tasks, effectively enabling information interaction and dependency modeling between intents and slots.

3.4. **Joint training.** This paper employs joint training to optimize and update the parameters of the two tasks of intent detection and slot filling. The formula for joint training is

$$L_1 \triangleq -\sum_{i=1}^{n}\sum_{j=1}^{N_I}\left(\hat{y}_i^{(j,I)}\log\left(y_i^{(j,I)}\right)+\left(1-\hat{y}_i^{(j,I)}\right)\log\left(1-y_i^{(j,I)}\right)\right) \tag{6}$$

$$L_2 \triangleq -\sum_{i=1}^{n}\sum_{j=1}^{N_S}\hat{y}_i^{(j,S)}\log\left(y_i^{(j,S)}\right) \tag{7}$$

where $\hat{y}_i^{(j,I)}$ and $\hat{y}_i^{(j,S)}$ are the ground truth labels for intent and slots, respectively, N_I is the number of intent labels, N_S is the number of slot labels, n is the number of words in the utterance, and the final joint loss formula is

$$L = \alpha L_1 + (1-\alpha)L_2 \tag{8}$$

where α is a hyperparameter, and the joint loss function can jointly optimize the two tasks to further reduce the error.

3.5. **ChatGPT.** This paper introduces a preprocessing module based on ChatGPT before the Shared Self-Attentive Encoder, utilizing the global semantic understanding capability of generative language models to clean, semantically normalize, and supplement potential information in the original input text, making it more suitable for subsequent intent recognition and slot filling tasks. Its workflow can be divided into the following two stages.

1. Text preprocessing and semantic context construction. In this stage, the system interacts the raw sentences with pre-set prompts, the content of which is "You are an expert in the field of spoken language understanding. Please do not add or remove punctuation marks or words. Without changing the order and number of words, correct the spelling in the following sentences and optimize colloquial words into more formal written expressions to make them clear in expression and grammatically correct for intent recognition and slot filling tasks". This prompt enforces the following strict constraints: ChatGPT

is prohibited from adding new words or removing existing ones, preserving the original sentence structure and information boundary. The sequence and total count of words remain unchanged, preventing unintended semantic drift. Under these constraints, the objectives of preprocessing include 1) Grammar and format adjustment: Automatically correct grammatical errors, irregular capitalization, tense confusion, etc., in the original sentences. 2) Removal of colloquial expressions: Optimize colloquial words in sentences to transform them into more formal written expressions. For example, change "talk, show, try" to more formal written forms such as "discuss, demonstrate, attempt". 3) Vocabulary and expression optimization: Refine vocabulary and clarify semantics for phenomena such as mixed use of synonyms and ambiguous expressions, ensuring that key intents and slot information are straightforward and easy to understand. 4) Potential information mining: Infer potentially overlooked implicit information based on the context without altering the original sentence structure, and apply it to text preprocessing to enhance the subsequent model's ability to parse semantics.

For example, the original sentence is "i gotta get the info and fix the bug before the meeting". After preprocessing by ChatGPT, the output is "I need get the information and fix the bug before the meeting". The advantages of this operation are as follows: eliminating colloquial expressions by replacing the colloquial "gotta" with the more formal written expression "need", making the sentence clearer; completing abbreviations by expanding "info" to "information", avoiding ambiguity and improving the model's parsability; improving capitalization by changing the lowercase "i" to uppercase "I" to conform to the norms of written language; without adding or deleting words, ensuring that the sentence is semantically more refined and the implied task information is more intuitive through vocabulary substitution and format adjustment.

2. Interactive semantic optimization and API invocation. By invoking the GPT-4 model API from OpenAI, ChatGPT performs grammar correction, expression refinement, and semantic completion on the input sentences according to the prompts, and the processed results are used as inputs for subsequent joint modeling. To ensure the stability of the overall process, the system has been extended and optimized in implementation. When invoking the API, if access failures or network anomalies occur, the system will automatically retry up to 5 times, with the waiting time increasing incrementally (e.g., 10 seconds for the first attempt, 20 seconds for the second, and 30 seconds for the third). This approach allows rapid recovery from temporary network fluctuations while avoiding resource waste caused by frequent retries. If multiple consecutive calls fail, the system triggers a circuit-breaking strategy, temporarily suspending calls to ChatGPT to prevent more severe failures or API rate limiting caused by frequent invalid requests. In addition, if the problem of connection difficulties is encountered, the difficulty of system integration can be reduced through offline batch preprocessing. The corpus can be pre-processed and the results cached to avoid real-time dependence on external services during the training stage.

This design fully leverages ChatGPT's strengths in cross-domain knowledge and deep semantic understanding during the preprocessing stage, providing more coherent and precise input expressions for subsequent multi-intent recognition and slot filling. It effectively reduces the risk of intent-slot conflicts caused by complex contexts and significantly enhances the system's generalization capability and overall performance in complex scenarios.

4. Experiments.

4.1. Datasets and metrics.

This study conducts experimental validation on two publicly available multi-intent utterance datasets, MixATIS and MixSNIPS [5,25,26]. The MixATIS dataset contains 13,162 training sentences, 756 validation sentences, and 828 test sentences; the MixSNIPS dataset contains 39,776 training sentences, 2,198 validation sentences, and 2,199 test sentences. In terms of evaluation metrics, this study follows the practices of previous research: using Slot F1 as the evaluation metric for the slot filling task, intent ACC as the evaluation metric for the multi-intent recognition task, and overall ACC as the evaluation metric for sentence-level semantic frame parsing. Among these, overall ACC measures the proportion of sentences in which both the intent and semantic slots are correctly predicted. The formula for calculating accuracy is as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

where TP is the true positive, FP is the false positive, FN is the false negative, and TN is the true negative.

The F1 score is the harmonic mean of precision and recall. Precision (P) is the ratio of true positives to the number of samples predicted as positive, and recall (R) is the ratio of true positives to the number of samples that are actually positive. The specific calculation formulas for precision, recall, and F1 score are as follows:

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{12}$$

4.2. Experimental settings.

In our experiments, the word embedding and hidden state are 256 in all datasets. The training batch size was set to 16, and the number of layers for all graph neural networks was set to 2. We used the Adam [27] optimizer for model training with a learning rate of 1e-3 and a weight decay of 1e-5. During training, we selected the model that performed best on the development set and reported its performance on the test set. All experiments were conducted on an RTX 4090.

4.3. Baselines.

We compared our model with the following baselines.

1) Stack-Propagation (Qin et al.) [28]: A joint model that proposes Stack-Propagation to capture intent semantic knowledge and perform token-level intent detection to further alleviate error propagation.

2) Joint Multiple ID-SF (Gangadharaiah) [3]: Employs a multi-task framework with a slot-gated mechanism for multi-intent recognition and slot filling.

3) AGIF (Qin et al.) [4]: Proposes an adaptive graph-interactive framework for joint multi-intent recognition and semantic slot filling, which extracts intent information for token-level slot filling.

4) GL-GIN (Qin et al.) [5]: Introduces a non-autoregressive method for joint multi-intent recognition and semantic slot filling, speeding up the model's decoding time and addressing the issue of incoherent semantic slots caused by non-autoregressive methods using local interaction graphs for semantic slots.

5) GISCo (Song et al.) [21]: A semantic understanding model based on constructing co-occurrence graphs from the entire corpus, achieving cross-task knowledge transfer by modeling the co-occurrence frequency relationship between semantic slots and intents.

6) SSRAN (Cheng et al.) [22]: Designs a Scope-Sensitive Result Attention Network based on the Transformer architecture, aiming to alleviate error propagation by capturing bidirectional interactions among results.

The details are shown in Table 1.

Table 1. Results comparison

Model	MixATIS			MixSNIPS		
	Intent (Acc/%)	Slot (F1/%)	Overall (Acc/%)	Intent (Acc/%)	Slot (F1/%)	Overall (Acc/%)
Stack-Propagation	72.1	87.8	40.1	96.0	94.2	72.9
Joint Multiple ID-SF	73.4	84.6	36.1	95.1	90.6	62.9
AGIF	74.4	86.7	40.8	95.1	94.2	74.2
GL-GIN	76.3	88.3	43.5	95.6	94.9	75.4
GISCo	75.0	88.5	48.2	95.5	95.0	75.9
SSRAN	77.9	89.4	48.9	98.4	95.8	77.5
Ours	78.4	89.6	49.9	98.6	96.2	78.3

4.4. **Main results.** This study proposes a joint modeling framework integrating graph attention networks and the large language model ChatGPT for multi-intent spoken language understanding tasks. Experimental results demonstrate that the proposed framework achieves significant performance improvements on both the MixATIS and MixSNIPS datasets. Specifically, on the MixATIS dataset, our model outperforms the baseline model SSRAN by 1.0%; on the MixSNIPS dataset, the improvement reaches 0.8%. These performance gains are attributed to three key innovations. First, the introduction of ChatGPT for comprehensive semantic optimization of raw text during preprocessing effectively enhances fine-grained intent recognition capabilities, particularly in resolving ambiguities between easily confused intents such as "Flight" and "Aircraft" in the aviation domain. ChatGPT injects domain-implicit knowledge to significantly improve the discriminability of semantic representations, addressing ambiguity issues in intent recognition caused by limited rules and domain knowledge in traditional methods. Second, the proposed bidirectional interaction mechanism between intents and semantic slots based on graph attention networks breaks through the limitations of traditional modeling. By modeling multi-type associations between intent nodes and slot nodes, the framework achieves a 1.0% overall accuracy improvement over the best baseline on the MixATIS dataset, validating the superiority of bidirectional interaction in multi-task collaboration. Third, the framework combines the bidirectional interaction mechanism with ChatGPT preprocessing, fully leveraging their synergistic advantages. Compared to unidirectional intent-to-slot guidance methods, the fusion of bidirectional interaction and ChatGPT not only leverages ChatGPT's deep semantic optimization of complex contexts during preprocessing but also utilizes global statistical priors captured by graph attention networks to effectively resolve ambiguity caused by similar labels in the aviation domain, thereby achieving greater gains in multi-intent detection tasks. These results validate the effectiveness of the bidirectional task-guidance mechanism and semantic enhancement strategies, and also demonstrate that the deep integration of large language models and traditional graph neural networks opens new possibilities for multi-intent understanding tasks.

4.5. **Model analysis.**

4.5.1. *Ablation test.* To study the effectiveness of each component in the model, we conducted ablation experiments on the MixATIS and MixSNIPS datasets.

Effectiveness of the ChatGPT module: To validate the critical role of the large language model ChatGPT in multi-intent joint modeling, this study systematically conducted ablation experiments on the MixATIS and MixSNIPS datasets. Specifically, we removed the ChatGPT preprocessing module from the framework and retained only the graph attention network for intent-slot interaction modeling (denoted as w/o ChatGPT). The experimental results are shown in Table 2. The data indicates that removing the ChatGPT module caused a drop in overall accuracy of 1.6% and 1.3% on the MixATIS and MixSNIPS datasets, respectively. This performance decline demonstrates that the finegrained semantic understanding provided by ChatGPT is an indispensable core component of the framework. Firstly, after removing ChatGPT, the global semantic consistency capture is insufficient. Graph attention networks are mainly good at mining the local structural relationships between nodes, but it is difficult to maintain the overall coherence and thematic consistency of discourse on a larger scale. The lack of a global information dissemination mechanism spanning multi-hop nodes makes it impossible to fully model the semantic connections of complex sentence patterns or long-distance dependencies. Secondly, the introduction of tacit knowledge and rules is limited. Pure graph models rely on explicit node and edge definitions. For intermediate tacit associations that do not appear or appear sparsely in the training set, such as industry term rules and common sense reasoning chains, they cannot be known, resulting in insufficient capture of subtle semantic differences, professional background information, and common sense constraints. Finally, the preprocessing module can dynamically adjust the input representation, introduce external knowledge or perform disambiguation according to the context. However, the pure graph model has fixed parameters during reasoning and lacks flexible coping strategies for the immediate changes of semantic combination patterns and sudden polysemous and ambiguous structures in the context, resulting in an increase in decision noise. To sum up, after removing the preprocessing module, the limitations of the graph attention network in handling global semantic coherence, tacit knowledge injection, cross-domain rule transfer, and dynamic context adaptation have become the main reasons for the performance degradation.

Table 2. Ablation results

Model	MixATIS			MixSNIPS		
	Intent (Acc/%)	Slot (F1/%)	Overall (Acc/%)	Intent (Acc/%)	Slot (F1/%)	Overall (Acc/%)
w/o Slot to Intent	76.9	88.6	47.2	97.7	95.5	77.2
w/o Intent to Slot	77.3	88.8	47.7	98.2	95.8	77.5
w/o ChatGPT	77.6	88.9	48.3	97.9	95.7	77.0
Ours	78.4	89.6	49.9	98.6	96.2	78.3

The effectiveness of the graph attention network-based bidirectional interaction mechanism: To validate the core value of the bidirectional interaction mechanism between intent and semantic slots based on graph attention networks, this study designs systematic ablation experiments: When removing the information transmission channel from slots to intents (denoted as w/o Slot to Intent), the model accuracy decreases by 2.7% and 1.1% on MixATIS and MixSNIPS, respectively; When removing the information transmission pathway from intent to slot (denoted as w/o Intent to Slot), the model accuracy decreases by 2.2% and 0.8% on MixATIS and MixSNIPS, respectively. This result demonstrates that intent recognition relies on the entity support provided by slot labels, while slot labels require dynamic adjustment through the domain knowledge framework formed by intent prediction. The absence of information flow in either direction of the bidirectional

interaction mechanism leads to significant performance degradation, and the synergistic optimization between intents and slots exhibits strong coupling – particularly in multi-intent scenarios, fine-grained semantics carried by slot labels (such as the flight number encoding rules in B-flight_number) provide critical supplementary clues for intent recognition. When the intention prediction fails to obtain these entity priors from the slot labels, the model's ability to distinguish complex intentions weakens accordingly. However, the global constraints at the intention level (such as the strong correlation between the "flight" intention and slots like airports and flight numbers) effectively enhance the consistency of slot labels. Once the intention information cannot be fed back to the slots, slot prediction is prone to semantic drift or conflicts. Without data complementarity in either direction, the model loses the ability to form a closed loop between explicit semantic prompts and implicit association reasoning, resulting in the breaking of the strong coupling relationship between the two. The semantics at each level cannot correct each other, and the overall reasoning efficiency and robustness decline accordingly. To sum up, the absence of information feedback in either direction will undermine the synergistic gain between intent and slots, weaken the dual guarantee of explicit entity support and global semantic constraints, thereby leading to a systematic decline in the performance of intention-slot joint modeling.

4.5.2. *Case study.* To further understand how ChatGPT works, we conducted a case study. When processing the sentence containing "What city is mco, how many first class flights does united have today and then what's restriction ap68", the system first preprocesses the sentence to meet the requirements of intent recognition and slot filling tasks. Without altering the word order and quantity, the preprocessing result is "What city is MCO, how many first class flights does United have today and then what's restriction AP68". The preprocessed sentence explicitly presents key information: airport code: MCO, cabin type: first class, airline: United, restriction code: AP68. Due to the existence of specific encoding rules in the aviation domain, the distinction between slots such as airport names, departure locations, and destinations is critically important. In the original sentence, "AP68" might be misclassified as a generic flight number. This error primarily stems from predefined rules lacking sufficient domain-specific implicit knowledge, leading to slot annotation confusion. By introducing ChatGPT for text preprocessing, the system leverages its implicit knowledge support for the aviation domain to identify that "AP68" should be recognized as an IATA code for specific restriction clauses in this context. In our approach, ChatGPT acts as a preprocessor to first perform deep semantic understanding and optimization of raw text, addressing the limitations of traditional methods in capturing fine-grained semantics. By incorporating open-domain knowledge, the model can not only identify implicit relationships in complex sentences but also effectively parse nested and lengthy expressions. This improvement not only enhances slot filling accuracy but also ensures logical coherence between intents and slots.

4.5.3. *Comparison of preprocessing effects based on different large language models.* To validate the superiority of using ChatGPT for preprocessing in this study, we also experimented with preprocessing methods based on other large models, as shown in Figure 5. Figure 5 demonstrates the accuracy rates achieved by different models in the joint modeling task of multi-intent recognition and slot filling. It can be observed that ChatGPT significantly outperforms DeepSeek and ERNIE Bot in this joint task. This is mainly attributed to ChatGPT's reliance on large-scale pretraining corpora and multi-layer attention mechanisms, which enable higher accuracy in parsing complex nested structures and long-range dependency relationships. Especially in addressing the "intent-slot segmentation" issue, its ability to optimize contextual coherence is particularly outstanding, for

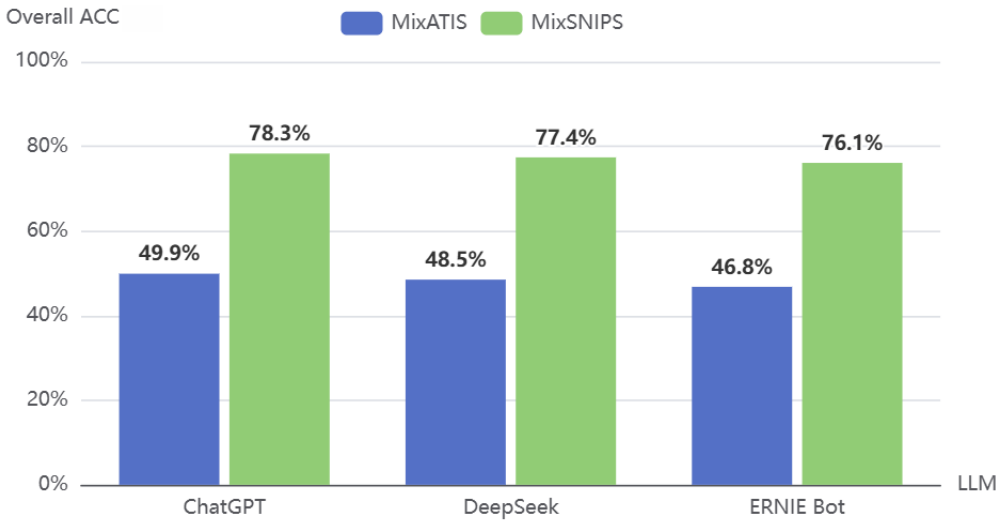

FIGURE 5. Comparison of different large language models

example, it can dynamically capture cross-sentence references and implicit logic, thereby effectively reducing semantic discontinuity. Additionally, ChatGPT's open-domain knowledge repository and generative architecture allow flexible adaptation to dynamic language environments. By actively exploring potential semantics rather than relying on static rules, it effectively compensates for the limitations of traditional methods in fine-grained ambiguity resolution. Finally, ChatGPT's end-to-end semantic reorganization during preprocessing significantly strengthens the implicit associations between intents and slots. In contrast, DeepSeek and ERNIE Bot exhibit relatively weaker performance in cross-task feature fusion due to differences in model capacity or training objectives. Experimental data conclusively demonstrate ChatGPT's indispensable superiority in handling complex semantic scenarios.

5. **Conclusion.** This paper proposes a joint modeling framework based on graph attention networks and the large language model ChatGPT for multi-intent spoken language understanding tasks. By constructing a bidirectional interaction mechanism between intents and semantic slots, the model effectively addresses the limitations of unidirectional guidance in traditional methods, achieving synergistic optimization of multi-intent detection and slot filling. Specifically, this work innovatively introduces ChatGPT to enhance dialogue context representation and utilizes graph attention networks to propagate information on two interaction graphs, fully exploring bidirectional dependencies between tasks. Experimental results demonstrate that the proposed method achieves outstanding performance on two public multi-intent datasets, validating the effectiveness of the bidirectional interaction mechanism and the ChatGPT-based semantic enhancement strategy. Future work will further explore more fine-grained semantic modeling methods to solve more complex multi-intent scenarios. Meanwhile, more detailed evaluation metrics will be introduced in the experimental evaluation to more comprehensively reveal the performance of the model under different sub-tasks and optimize the computational efficiency, so as to promote its practical application in real dialogue systems.

Acknowledgment. This work was supported by the Scientific Research Fund for the Higher Education Institutions of Liaoning Province of China under Grant LJ212410152070.

REFERENCES

[1] S. Young, M. Gašić, B. Thomson and J. D. Williams, POMDP-based statistical spoken dialog systems: A review, *Proceedings of the IEEE*, vol.101, no.5, pp.1160-1179, 2013.

[2] G. Tur and R. De Mori, *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*, John Wiley and Sons, 2011.

[3] R. Gangadharaiah, Joint multiple intent detection and slot labeling for goal-oriented dialog, *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019.

[4] L. Qin, X. Xu, W. Che and T. Liu, AGIF: An adaptive graph-interactive framework for joint multiple intent detection and slot filling, *arXiv Preprint*, arXiv: 2004.10087, 2020.

[5] L. Qin, F. Wei, T. Xie, X. Xu, W. Che and T. Liu, GL-GIN: Fast and accurate non-autoregressive model for joint multiple intent detection and slot filling, *arXiv Preprint*, arXiv: 2106.01925, 2021.

[6] B. Xing and I. W. Tsang, Group is better than individual: Exploiting label topologies and label relations for joint multiple intent detection and slot filling, *arXiv Preprint*, arXiv: 2210.10369, 2022.

[7] P. Veličković, G. Cucurull, A. Casanova et al., Graph attention networks, *arXiv Preprint*, arXiv: 1710.10903, 2017.

[8] D. Chen, L. O'Bray and K. Borgwardt, Structure-aware transformer for graph representation learning, *Proc. of the 39th International Conference on Machine Learning*, pp.3469-3489, 2022.

[9] P. Niu, Z. Chen and M. Song, A novel bi-directional interrelated model for joint intent detection and slot filling, *arXiv Preprint*, arXiv: 1907.00390, 2019.

[10] Z. Zhu, X. Cheng, H. An, Z. Wang, D. Chen and Z. Huang, Zero-shot spoken language understanding via large language models: A preliminary study, *Proc. of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp.17877-17883, 2024.

[11] H. Fong and E. Ong, Evaluating ChatGPT for joint intent detection and slot filling: Zero-shot vs. few-shot prompting, *PCSC*, 2023.

[12] B. Liu and I. Lane, Attention-based recurrent neural network models for joint intent detection and slot filling, *arXiv Preprint*, arXiv: 1609.01454, 2016.

[13] C. W. Goo, G. Gao, Y. K. Hsu, C. L. Huo, T. C. Chen, K. W. Hsu and Y. N. Chen, Slot-gated modeling for joint slot filling and intent prediction, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol.2, pp.753-757, 2018.

[14] C. Li, L. Li and J. Qi, A self-attentive model with gate mechanism for spoken language understanding, *Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp.3824-3833, 2018.

[15] H. E, P. Niu, Z. Chen and M. Song, A novel bi-directional interrelated model for joint intent detection and slot filling, *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019.

[16] B. Kim, S. Ryu and G. G. Lee, Two-stage multi-intent detection for spoken language understanding, *Multimedia Tools and Applications*, vol.76, pp.11377-11390, 2017.

[17] Y. Wang, Y. Shen and H. Jin, A bi-model based RNN semantic frame parsing model for intent detection and slot filling, *arXiv Preprint*, arXiv: 1812.10235, 2018.

[18] Z. Zhang, L. Ma, D. Zhang, X. Yan and H. Wang, Graph LSTM with context-gated mechanism for spoken language understanding, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol.34, no.5, pp.9539-9546, 2020.

[19] C. Chen, P. Zhou and Y. Zou, Joint multiple intent detection and slot filling via self-distillation, *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2022)*, pp.7612-7616, 2022.

[20] C. Chen, N. Chen, Y. Zou, Y. Wang and X. Sun, A transformer-based threshold-free framework for multi-intent NLU, *Proc. of the 29th International Conference on Computational Linguistics*, pp.7187-7192, 2022.

[21] M. Song, B. Yu, Q. Li, Y. Wang, T. Liu and H. Xu, Enhancing joint multiple intent detection and slot filling with global intent-slot co-occurrence, *Proc. of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp.7967-7977, 2022.

[22] L. Cheng, W. Yang and W. Jia, A scope sensitive and result attentive model for multi-intent spoken language understanding, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol.37, no.11, pp.12691-12699, 2023.

[23] A. Abboud and G. Oz, Towards equitable natural language understanding systems for dialectal cohorts: Debiasing training data, *Proc. of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp.16487-16499, 2024.
[24] A. Joshi, R. Dabre, D. Kanojia, Z. Li, H. Zhan, G. Haffari and D. Dippold, Natural language processing for dialects of a language: A survey, *ACM Computing Surveys*, vol.57, no.6, pp.1-37, 2025.
[25] C. T. Hemphill, J. J. Godfrey and G. R. Doddington, The ATIS spoken language systems pilot corpus, *Proc. of the Workshop on Speech and Natural Language*, pp.96-101, 1990.
[26] A. Coucke, A. Saade, A. Ball, T. Bluche, A. Caulier, D. Leroy and J. Dureau, Snips voice platform: An embedded spoken language understanding system for private-by-design voice interfaces, *arXiv Preprint*, arXiv: 1805.10190, 2018.
[27] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv Preprint*, arXiv: 1412.6980, 2014.
[28] L. Qin, W. Che, Y. Li, H. Wen and T. Liu, A stack-propagation framework with token-level intent detection for spoken language understanding, *arXiv Preprint*, arXiv: 1909.02188, 2019.

Author Biography

Shuxin Chen received her bachelor's degree from the School of Data Science and Computer Science of Shandong Women's College, China, in 2023. She is currently a Master's student in School of Information Science and Engineering at Dalian Polytechnic University, China, Class of 2023. Her primary research interests are natural language processing and artificial intelligence.

Xu Li received B.S. degree in Computer Science from University of Science and Technology Anshan, China, in 2003. She received M.E. and Ph.D. degrees in Computer Application Technology from Yanshan University, China, in 2006 and 2010, respectively. She is currently an associate professor in the Innovation and Entrepreneurship Education Center, Dalian Polytechnic University, China. Her current research interests include natural language processing and deep learning.

Jiaqi Wang received his bachelor's degree from the School of Intelligent Engineering, Jinan University Quancheng College, China, in 2022. He is currently a Master's student in School of Information Science and Engineering at Dalian Polytechnic University, China, Class of 2023. His primary research interests are natural language processing and federal learning.

Yu Zhang received her bachelor's degree from the School of Computer Science and Engineering at Yantai University, China, in 2023. She is currently a Master's student in School of Information Science and Engineering at Dalian Polytechnic University, China, Class of 2023. Her primary research interests are natural language processing and artificial intelligence.