

# 基于图注意力网络的多意图识别与语义槽填充联合模型\*

张玉慧, 陈黎, 琚生根, 李每文



(四川大学 计算机学院, 四川 成都 610065)

通信作者: 琚生根, E-mail: [jsg@scu.edu.cn](mailto:jsg@scu.edu.cn)

**摘要:** 口语理解是任务型对话系统的关键任务, 主要由语义槽填充和意图识别两个子任务组成. 目前主流的方法是对语义槽填充和意图识别进行联合建模. 虽然这种方法在语义槽填充和意图识别上都取得了不错的效果, 但依然存在联合建模中意图识别和语义槽填充交互过程的错误传播问题以及多意图场景下多意图信息与语义槽信息的错误对应问题. 针对上述问题, 提出一种基于图注意力网络的多意图识别与语义槽填充联合模型 (WISM). WISM 模型通过细粒度 (单词级) 意图与语义槽建立单词级别一对一映射关系以修正多意图信息与语义槽之间的错误对应关系, 然后通过构建单词-意图-语义槽的交互图, 并利用细粒度下的图注意力网络建立两个任务之间的双向联系以此来降低交互过程中错误传播问题. 在 MixSINPS 和 MixATIS 数据集上的实验结果表明, WISM 相较于现有的最新模型在语义准确率分别提高 2.58% 和 3.53%. 所提模型在提高语义准确率的同时展示了多意图信息与语义槽之间的映射关系.

**关键词:** 口语理解; 多意图识别; 语义槽填充; 联合建模; 图注意力网络

**中图法分类号:** TP18

中文引用格式: 张玉慧, 陈黎, 琚生根, 李每文. 基于图注意力网络的多意图识别与语义槽填充联合模型. 软件学报, 2024, 35(12): 5509-5525. <http://www.jos.org.cn/1000-9825/7062.htm>

英文引用格式: Zhang YH, Chen L, Ju SG, Li MW. Joint Model for Multi-intent Detection and Slot Filling Based on Graph Attention Network. Ruan Jian Xue Bao/Journal of Software, 2024, 35(12): 5509-5525 (in Chinese). <http://www.jos.org.cn/1000-9825/7062.htm>

## Joint Model for Multi-intent Detection and Slot Filling Based on Graph Attention Network

ZHANG Yu-Hui, CHEN Li, JU Sheng-Gen, LI Mei-Wen

(College of Computer Science, Sichuan University, Chengdu 610065, China)

**Abstract:** Spoken language understanding is a key task in task-based dialogue systems, mainly composed of two sub-tasks: slot filling and intent detection. Currently, the mainstream method is to jointly model slot filling and intent detection. Although this method has achieved good results in both slot filling and intent detection, there are still issues with error propagation in the interaction process between intent detection and slot filling in joint modeling, as well as the incorrect correspondence between multi-intent information and slot information in multi-intent scenarios. In response to these problems, this study proposes a joint model for multi-intent detection and slot filling based on graph attention networks (WISM). The WISM established a word-level one-to-one mapping relationship between fine-grained intentions and slots to correct incorrect correspondence between multi-intent information and slots. By constructing an interaction graph of word-intent-semantic slots and utilizing a fine-grained graph attention network to establish bidirectional connections between the two tasks, the problem of error propagation during the interaction process can be reduced. Experimental results on the MixSINPS and MixATIS datasets showed that, compared with the latest existing models, WISM has improved semantic accuracy by 2.58% and 3.53%, respectively. This model not only improves accuracy but also verifies the one-to-one correspondence between multi-intent and semantic slots.

**Key words:** spoken language understanding; multi-intent detection; slot filling; joint modeling; graph attention network (GAT)

\* 基金项目: 国家自然科学基金重点项目 (62137001); 四川省重点研发项目 (2023YFG0265)  
收稿时间: 2023-06-26; 修改时间: 2023-08-05; 采用时间: 2023-09-03; jos 在线出版时间: 2024-02-05  
CNKI 网络首发时间: 2024-02-09

口语理解是对话系统中的关键模块,它将用户输入的自然语言转化为机器能够理解的语义表达。口语理解包含意图识别(intent detection, ID)与语义槽填充(slot filling, SF)两个任务。其中,意图识别也叫意图分类,是分析理解输入句子中用户的行为,并为其分配一个或多个意图标签,通常归结为分类问题,而槽填充任务则是对句子中有意义的单词进行标记,并将其与预定义的槽标签对应起来,通常被视为序列标注任务<sup>[1]</sup>。因此,这两个子任务的性能直接影响到口语理解的准确性和对话系统的表现效果。

在对话理解的早期研究中,将意图识别与语义槽填充任务被视为独立的,采用分别建模的方法。这种方法忽略了两个任务之间的语义关联,也没有充分利用两个任务中的信息来相互促进<sup>[1]</sup>。因此,现有的研究主要关注如何将两个任务进行联合建模。早期的联合建模主要采用共享模型参数的隐式建模方式<sup>[2,3]</sup>,这类模型虽然性能优于独立建模,但仍然忽略了两个任务之间的联系性<sup>[4]</sup>,例如意图信息对语义槽预测具有约束作用,而语义槽信息对意图识别任务起到辅助作用。近年来,越来越多的研究者对两个任务之间的关联进行显式建模研究。例如, Goo 等人<sup>[5]</sup>和 Li 等人<sup>[6]</sup>采用门控机制实现意图信息整合到语义槽任务中的信息流,提高了意图识别与语义槽填充两个任务的性能,但是上述模型仅考虑意图到语义槽任务的单向联系性,未考虑语义槽信息对于意图识别任务的辅助作用。为此, E 等人<sup>[7]</sup>提出了双向相关的模型来完成联系建模,并采用迭代机制进一步加强意图与语义槽之间的相互关联性。

现有的研究利用门控机制、迭代机制、胶囊网络<sup>[8]</sup>、记忆网络<sup>[9]</sup>、图神经网络<sup>[10]</sup>等方法对意图识别和语义槽填充任务进行联合建模。例如 Zhang 等人<sup>[8]</sup>提出了一种基于胶囊的神经网络模型,通过动态的协议路由完成语义槽填充和意图识别,利用推断的意图表示进一步协同语义槽填充性能,提高模型的整体性能。随着图神经网络进入到人们的视野,研究人员逐渐将图神经网络应用到模型中,例如: Tang 等人<sup>[10]</sup>构建一个基于图的 CRF 图卷积网络(GCN)来学习单词、插槽标签和意图标签之间的关系; Wei 等人<sup>[11]</sup>提出了一种新的轮图注意网络(wheel-GAT)联合模型,直接建模相关连接,用于单意图识别和语义槽填充。

现有意图识别与语义槽联合模型主要关注于单意图场景<sup>[12]</sup>,然而,在实际场景中,用户输入的语句可能包含多个意图<sup>[13,14]</sup>。因此,研究者开始借鉴多任务学习的思想来构建多意图识别与槽填充联合框架。其中, Gangadharaiah 等人<sup>[15]</sup>首次提出多任务框架下的多意图识别与槽填充的联合框架,将多意图识别任务看作多标签分类任务,针对多意图识别任务提出了采用句子级和单词级的联合多意图识别,但在建立多意图与语义槽之间联系时使用句子级的多意图信息通过 slot gate 与语义槽之间建立联系,但是门控机制不能充分总结和记录意图信息。为此, Qin 等人<sup>[16]</sup>提出了一种联合多意图识别和语义槽填充的自适应图交互框架,采用图神经网络架构实现意图与语义槽之间的强关联性;类似地,2021年, Qin 等人<sup>[17]</sup>再次使用图神经网络构建局部语义槽交互图和全局意图-语义槽交互图用以建模两个任务之间的联系性。虽然上述模型采用图神经网络缓解了门控机制在意图信息总结和记录方面的缺点,但是上述模型在意图识别与语义槽填充联系建模过程中,将所有意图信息与语义槽进行联系,产生了错误的对应关系,并且两个任务交互过程中采用预测后的意图信息对语义槽填充任务进行指导,会产生错误传播问题。

无论是在单意图场景还是在多意图场景,意图与语义槽之间多呈现为一对多关系。在单意图场景中,同一语句中的所有语义槽共享同一个意图;然而在多意图场景中,同一用户语句中不同语义槽可能对应着不同的意图。如图 1(b)中“ramy ayach”和“latin pop rising”语义槽对应“AddToPlaylist”意图,“schedule”和“the Kentuckians”语义槽对应意图“SearchScreeningEvent”。然而,现有大多数多意图识别和语义槽联合模型<sup>[15-17]</sup>要么沿用单意图场景下的思想,即所有的语义槽共享多意图信息整合后的一个上下文信息,要么将所有的语义槽同时建模联系该句中所有意图,如图 1(a)所示。当句子中的一个语义槽与多个意图之间建立关系<sup>[16,17]</sup>,会导致一种错误的对应关系,同时联合模型中在意图识别与语义槽填充两个任务进行交互的过程会产生错误传播,即当某一个任务产生错误的预测信息会影响另一任务的预测过程。

针对上述问题,本文提出了一种基于图注意力网络的多意图识别与语义槽填充联合模型(word-intent-slot-mapping, WISM)。WISM 模型通过将意图建模到单词级别,从而构建了一种细粒度的单词-意图-语义槽映射关系。此外, WISM 模型通过构建单词-意图-语义槽交互图来实现意图与语义槽任务之间的双向联系,并使用图注意力网络来捕获三者之间的交叉影响。在 MixSINPS 和 MixATIS 数据集上进行实验验证, WISM 模型证明了其在语义准确性上表现出了良好的性能。

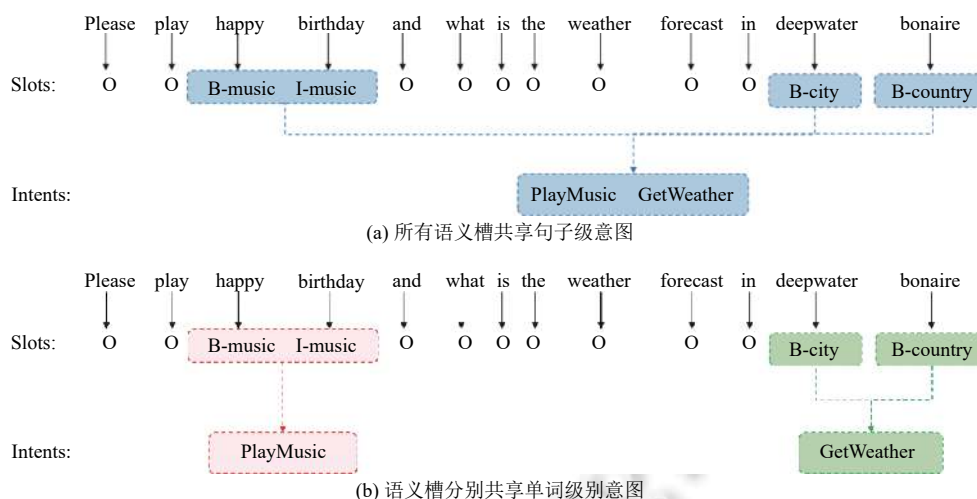


图1 多意图场景下意图与槽联系方式

本文的主要贡献可归纳如下。

(1) 为了解决多意图识别和槽填充联合模型中两个任务交互中存在的错误传播问题和错误对应关系, 本文提出了 WISM 模型。

(2) 通过构建单词-意图-语义槽图交互层, 利用图注意力网络进行单词级别的意图与语义槽之间双向建模, 充分利用细粒度下两个任务之间的交叉影响, 降低交互过程中的错误传播问题。

(3) 利用单词级意图产生与语义槽之间的单词级别一对一映射关系, 避免多意图信息与语义槽信息之间的错误映射关系。

(4) 通过修正多意图与语义槽错误对应关系和降低交互过程中的错误传播, 通过实验验证表明, 本文提出的 WISM 模型在 MixSINPS 和 MixATIS 数据集上提高了语义准确度。

本文的第 1 节介绍意图识别与语义槽填充联合模型的相关方法和研究现状。第 2 节对意图识别与语义槽填充联合任务进行形式化问题定义。第 3 节介绍本文构建的基于图注意力网络的多意图与语义槽联合模型 WISM。第 4 节介绍实验数据集和评估指标, 并进行了对比实验、消融实验以及可视化分析验证所提模型的有效性。最后, 第 5 节总结全文, 并展望未来的研究工作。

## 1 相关工作

口语理解中包含意图识别与语义槽填充两个任务, 由于两个任务同时出现, 随着研究的深入, 研究人员发现两个任务之间能够相互促进, 因此人们常通过各种方式对两个任务进行联合建模。

随着深度学习的不断学习及深入, 近年来, 基于深度学习的意图识别与语义槽填充联合模型取得了不错的效果。最初提出的意图识别与语义槽填充联合模型采用共享编码器来捕捉公共特征的隐式建模方式。Guo 等人<sup>[18]</sup>提出使用递归神经网络 (RecNN) 建模意图识别与语义槽填充联合模型, 模型采用树结构, 叶子节点表示单词, 根节点的输出向量用以意图识别, 前一路径、后一路径、当前路径三路径向量用以语义槽填充。为简化树形结构与消除由于语法树生成的时间和人工等成本, 2016 年, Zhang 等人<sup>[3]</sup>提出利用门控循环单元 GRU (gated recurrent unit) 学习每个单词的语义表示信息, 从而预测每个单词的语义槽标签, 同时使用最大池化层 (max-pooling layer) 来获取句子的全局语义表示, 用以预测句子的意图识别结果, 最后使用同一的损失函数进行联合训练, 并取得了不错的效果。相似的, Zhou 等人<sup>[19]</sup>使用两层 LSTM 进行联合建模, 第 1 层实现语义槽填充任务, 第 2 层实现意图识别任务, 最后使用联合损失函数; 此外, Liu 等人<sup>[20]</sup>提出了一个基于 RNN 的联合模型, 并使用共享的 RNN 来进行意图识别和语义槽填充的建模。上述各种模型虽然实现了联合建模, 相较于独立模型而言, 在意图识别与语义槽填充两个任

务上均有较大的性能提升;但是上述模型在共享编码层部分多使用 RNN 及其变种 (LSTM、GRU) 处理序列数据信息,内部结构中存在着循环结构,难以进行并行化计算,并且忽略了两个任务之间的联系性,及意图识别预测的结果往往能够指导语义槽填充任务的预测。

由于注意力机制<sup>[21]</sup>能够学习句子的上下文表示,并且能够并行化计算,研究人员探索将注意力机制应用到意图识别与语义槽填充联合模型的编码层中<sup>[5,6,22]</sup>。随着门控机制 (gate mechanism)/掩码机制 (mask mechanism) 在信息共享上的广泛应用,研究人员逐渐将门控机制/掩码机制应用在意图识别与语义槽填充联合建模中<sup>[5,6]</sup>,利用上述机制实现两个任务之间的信息交互。Li 等人<sup>[6]</sup>提出了一种具有门控机制的新型自注意模型,以充分利用槽位和意图之间的语义相关性;模型首先基于注意力机制获取意图增强的嵌入,然后首次利用意图增强的门控机制来匹配语义槽填充的标签,最终通过端到端的联合学习同时优化两个任务的目标。同样的, Goo 等人<sup>[5]</sup>提出使用门控机制来模拟意图识别与语义槽填充之间的关系,模型使用门控机制学习意图与语义槽注意力向量之间的关系,通过全局优化得到更好的语义框架。由于采用门控机制来控制意图信息,不能充分的总结和记录意图信息, Qin 等人<sup>[23]</sup>提出采用一种堆栈传播框架,直接将意图识别的输出作为语义槽填充任务的输入,直接利用了意图的信息来指导语义槽填充任务。上述模型利用门控机制或堆栈传播框架使用意图信息增强语义槽填充任务预测的单向信息传递,在一定程度上促使了两个任务性能的提高;但语义槽填充信息对意图任务之间也有着一定的促进作用,及两个任务之间有着双向的关联性。

为了建立两个任务之间的双向联系性<sup>[24]</sup>, Soyeon 等人<sup>[25]</sup>提出用于意图识别与语义槽填充双向联合模型,模型首先通过 BERT 对用户输入语句进行嵌入表示,随后提出包含用于意图识别的语义槽到意图 (intent2slot) 和用于语义槽填充的意图到语义槽 (slot2intent) 两个模型的双向口语理解层,其中语义槽到意图模型是将语义槽任务中的序列概率分布用作意图识别任务中的附加条件,帮助意图识别任务的预测,同样的意图到语义槽模型旨在利用整个语句的意图标签概率来辅助语义槽填充任务的预测。相同地,文献<sup>[7]</sup>提出 SF-ID 网络建立两个任务之间的直接联系, SF-ID 网络中有一个 SF 子网和一个 ID 子网组成,通过定义两个子网的顺序实现前一个子网通过引入相关因子实现对后一个子网的积极作用,然后通过迭代机制进行训练,进一步实现两个任务之间的相互促进关系。

随着深度学习技术的不断发展,研究人员逐渐将图神经网络应用到意图识别与语义槽填充联系训练中,显式表现两个任务之间关系的同时学习到单词与标签之间的全局关系。例如,2020年,Zhang 等人<sup>[26]</sup>提出 S-LSTM 图神经网络,采用通过的 LSTM 单词表示和句子表示作为节点,处理整个句子的同时进行全局的协调;同样的 Tang 等人<sup>[10]</sup>构建了一个基于图的 CRF 图卷积网络 (GCN),用以学习单词、插槽标签和意图标签之间的关系;2021年, Tao 等人<sup>[27]</sup>结合句法信息使用 GCN 训练句子的依赖解析,以单词作为节点,标记的依赖作为边,从而得到最终的结果。2023年, Wu 等人<sup>[28]</sup>提出了一种新的图卷积 LSTM (GC-LSTM) 编码器,通过合并一个强大的图卷积算子来学习对话依赖图所包含的语义,不仅可以捕获对话中的时空语义特征,还可以学习意图识别和语义槽填充之间的共现关系。上述阐述模型以单意图为出发点建立意图识别与语义槽填充联合模型,但是在现实场景中,用户输入的一句话中可能存在着多个意图,为此,研究人员将意图识别与语义槽填充联合建模从单意图场景转移到多意图场景。Gangadharaiah 等人<sup>[15]</sup>将多意图识别任务看作多标签分类任务,依旧采用门控机制实现意图信息到语义槽任务的指导作用,但是模型中直接采用句子级意图信息与语义槽任务建立联系。同样地, Qin 等人<sup>[17]</sup>提出了一种全局-局部交互网络,采用局部语义槽感知层建立语义槽之间的依赖性以缓解语义槽不均衡问题,全局意图-语义槽交互层建模话语中多意图信息与语义槽之间的联系性;但是在建立多意图与语义槽之间的联系性时,模型将多意图信息与全部的语义槽进行联系,导致意图与语义槽信息之间的对应关系产生问题,同时在意图与语义槽交互过程中,一旦其中一个任务出现错误的预测结果,极有可能影响另一任务的预测结果,即两个任务之间交互过程存在错误传播问题。

从上述研究可知,多意图识别与语义槽填充联合模型上存在着两个主要问题。

(1) 联合建模过程中,意图和语义槽交互过程的错误传播问题,即交互过程中意图结果错误是会对模型引入错误意图信息,误导语义槽的结果预测,反之也一样。

(2) 意图对于语义槽存在着一对多的映射关系,在单意图场景中,所有语义槽共享同一意图;但在多意图场景

中,不同的语义槽对应着不同的意图.无论直接采用句子级的意图对语义槽进行指导,还是将句子中所有的意图与语义槽建立联系,均会产生意图与语义槽之间的错误对应关系.

## 2 问题定义

给定一个自然语言句子  $X = \{x_1, x_2, \dots, x_n\}$ ,  $x_i$  表示输入语句中的单词,多意图识别可以表述为一个多标签分类任务,输出多个与输入话语对应的意图标签  $O^I = \{o_1^I, o_2^I, \dots, o_m^I\}$ .槽填充是将每个  $x_i$  映射到槽标签的序列标记任务,输出语义槽标签序列  $O^S = \{o_1^S, o_2^S, \dots, o_n^S\}$ ,其中  $n$  表示输入语句  $U$  的单词长度,  $m$  表示语句  $X$  对应的多个意图的个数.给定用户输入语句的口语理解实例如表 1 所示.

表 1 口语理解实例

Words	Please	play	happy	birthday	and	what	is	the	weather	forecast	in	deepwater	bonaire
Slots	O	O	B-music	I-music	O	O	O	O	O	O	O	B-city	B-country
Intents	—	—	PlayMusic		—	—	—	—	—	—	—	GetWeather	

为了保证单词级意图识别在多意图场景下的准确运用,我们提出了下列假设.

(1) 将单词级意图任务看作单词级多分类任务,单词的意图作为分类结果,对于某些没有意图的单词,采用特殊的标签“O”作为该单词的意图,因而每个单词均有所属的意图且仅属于单一意图.

(2) 除去“O”单词级的意图,同一输入语句的单词级意图与句子级意图属于包含关系,即  $o_i^I \in o^I$ ,且同一句话中的所有单词级意图是句子级意图的子集,即  $\sum o_i^I \subset o^I$ , $o_i^I$  表示语句中的单词级意图, $o^I$  表示该语句中的句子级意图.

(3) 同一输入语句的单词级意图与语义槽属于一一对应关系,单词级意图“O”对应语义槽的 O 标签,其他的单词级意图对应于语义槽的 B-name 和 I-name 标签.

## 3 本文方法

针对多意图与语义槽错误对应关系和交互过程中的错误传播问题,本文提出 WISM 模型,首先利用单词级意图识别构建意图与语义槽之间一对一映射关系,解决多意图场景下多意图与语义槽信息之间的错误对应关系,然后构建单词-意图-语义槽交互层,通过图注意力网络建模 token 级别的多意图与语义槽之间的联系,减轻交互过程中的错误传播问题.WISM 框架如后文图 2 所示.

WISM 模型包括初始层、图交互层、预测层:初始层是通过 BERT 对上下文进行编码,并通过意图解码器与语义槽解码器获取初始的单词级意图结果、句子级意图结果和语义槽序列;图交互层是构建单词-意图-语义槽交互图,通过图注意力机制计算单词、意图和语义槽三者之间的联系性;预测层是将图交互层的结果分别输入到不同的 LSTM 中,然后对 LSTM 的输出结果进行解码操作,得到最终的多意图识别结果和语义槽序列.下面将对模型的各层给出详细的介绍.

### 3.1 初始层

初始层包括编码器和 3 种解码器(句子级意图解码器、单词级意图解码器、语义槽解码器):编码器是将词向量转化为用户输入语句的语义隐藏向量,句子级意图解码器是通过特殊标记 [CLS] 的语义隐藏向量获取句子对应的意图,单词级意图解码器是通过每个单词的语义隐藏向量获取每个单词的意图,然后通过投票机制获取整个句子的意图,语义槽解码器是通过每个单词的语义隐藏向量获取每个单词对应的语义槽.

#### (1) 编码器

模型使用 BERT<sup>[29]</sup> 作为模型 token 序列的编码层,前面加入特殊的标记 [CLS],后面加入 [SEP],句子  $X = \{\text{[CLS]}, x_1, x_2, \dots, x_n, \text{[SEP]}\}$  作为输入,输出为  $H = \{h_{\text{CLS}}, h_1, h_2, \dots, h_n, h_{\text{SEP}}\}$  将用于后续工作,其中  $h_i \in \mathbb{R}^d$ .

$$H = \text{BERT}(X) \quad (1)$$

其中,  $x_i$  为  $u_i$  的词向量,通过 BERT 获得用户语句的隐藏状态  $h_i$ .

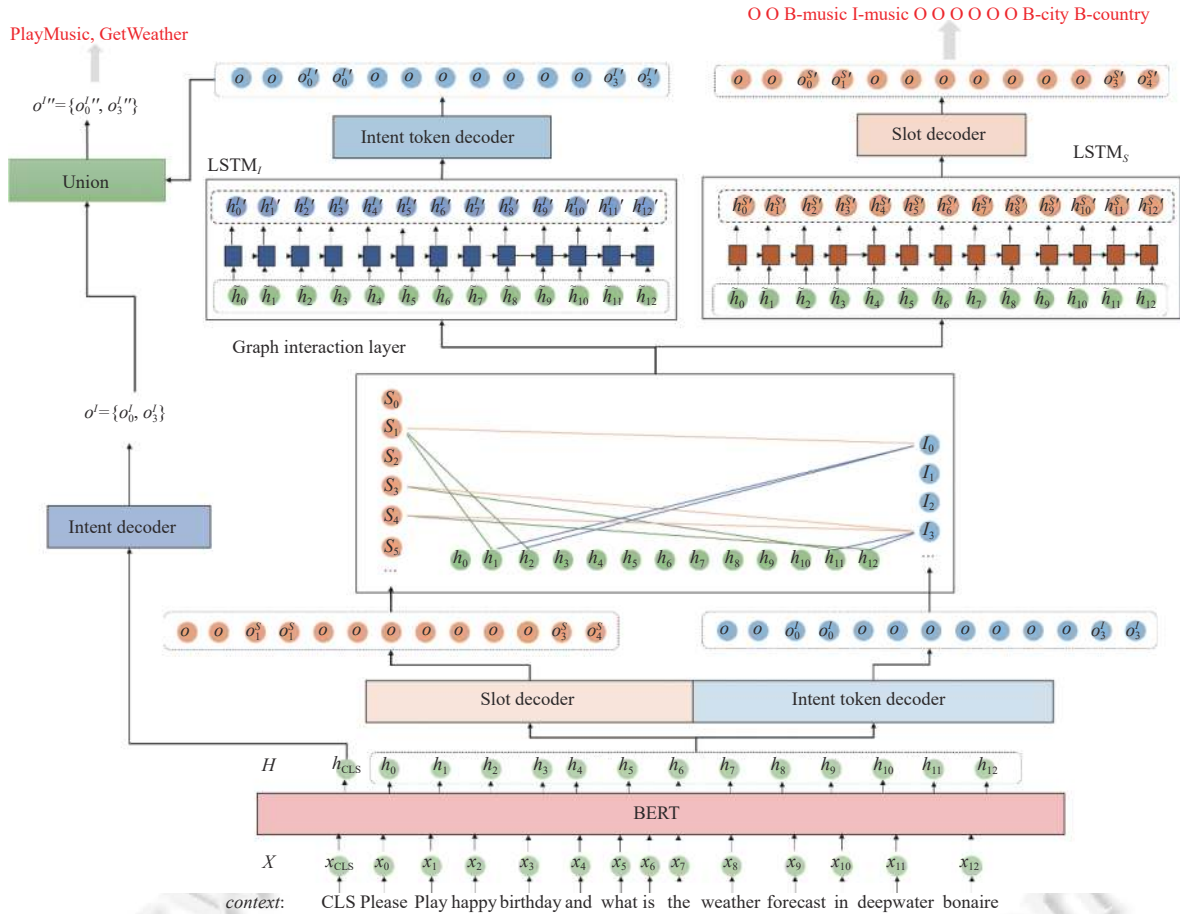


图 2 WISM 模型结构

(2) 句子级意图识别 (intent decoder)

根据文献 [15], 本文将多意图识别任务看作多标签分类任务, 采用单词级和句子级两个级别同时对多意图进行识别, 随后将两个级别的意图识别结果进行融合得到最终意图识别结果.

为了对句子级别的意图  $y^l$  进行分类, 首先使用 Sigmoid 函数作为编码层输出  $h_{CLS}$  的激活函数, 公式如下所示:

$$y^l = \sigma(W^l h_{CLS} + b^l) \tag{2}$$

其中,  $W^l \in \mathbb{R}^{l \times d}$ ,  $b^l \in \mathbb{R}^l$  为可训练参数,  $y^l$  表示每个意图标签的概率.

在推测阶段, 我们使用阈值超参  $\delta$  和意图标签的概率  $y^l = \{y^l_1, y^l_2, \dots, y^l_{|l|}\}$  预测句子的意图  $o^l = \{o^l_1, o^l_2, \dots, o^l_n\}$ , 其中  $0 < \delta < 1$ ,  $o^l$  表示比  $\delta$  高的概率  $y^l$ . 例如  $y^l = \{0.3, 0.7, 0.6, 0.1\}$ ,  $\delta = 0.5$ , 此时我们预测的句子级别的意图识别结果是  $o^l = \{o^l_1, o^l_2\}$ .

(3) 单词级意图识别 (intent token decoder)

单词级别的意图识别首先预测每个单词上的意图, 然后通过使用投票机制对所有单词进行意图投票获得意图识别的结果.

使用编码层的输出  $h_i$  用作单词级别的意图识别, 公式如下所示:

$$y^l_i = \sigma(W^l_i (LeakyReLU(W_i h_i + b_i) + b^l_i)) \tag{3}$$

其中,  $W^l_i, W_i$  为可训练参数,  $\sigma$  表示 Sigmoid 函数,  $y^l_i$  表示第  $i$  个单词的意图结果.

最后通过单词级别意图获取的句子意图结果方式如下所示:

$$I_i = \begin{cases} 1, & \text{if } I_{(i,k)} > 0.5 \\ 0, & \text{others} \end{cases} \quad (4)$$

$$o^j = o_k^j \left( \sum_{i=1}^n I_i \right) > \frac{n}{2} \quad (5)$$

其中,  $I_{(i,k)}$  表示第  $i$  个单词对于第  $k$  个意图标签的概率.

当意图标签在句子中  $n$  个单词中得到超过一半的正面预测 (肯定票) 时, 本文将其预测为话语的意图. 例如: 如果一句话中的语义表示为  $X = \{x_1, x_2, x_3\}$ , 即句子中由 3 个单词组成, 该句的候选意图标签有  $\{o'_0, o'_1, o'_2, o'_3\}$  这 4 个标签, 预测得知每个单词对应的意图标签概率分别是:  $I_1 = \{0.1, 0.4, 0.7, 0.9\}$ ,  $I_2 = \{0.6, 0.3, 0.6, 0.8\}$  和  $I_3 = \{0.2, 0.3, 0.9, 0.8\}$  若概率大于 0.5 设置为投肯定票 (正面预测), 则可以通过 3 个单词的意图识别的概率, 得到  $\{1, 0, 3, 3\}$  意图的肯定票数, 因此获取半数 (3 个单词即为 3/2) 以上的意图标签是  $\{o'_2, o'_3\}$ , 从而可以得到通过单词级意图识别方法获得的意图识别结果为  $o^j = \{o'_2, o'_3\}$ .

#### (4) 语义槽填充解码器 (slot decoder)

大部分研究人员将语义槽填充任务看作序列标注任务, 采用 CRF 进行语义槽结果预测, 但是其运行时间较长, 为了提高运行速度, 本文采用非自回归方法用来提升整体模型的速度, 将语义槽填充任务看作单词级别的多分类任务, 使用编码层的输出  $h_i$  作为语义槽填充任务的输入, 公式如下所示:

$$y_i^s = \sigma(W_i^s (\text{ReLU}(W_s h_i + b_s) + b_i^s)) \quad (6)$$

其中,  $W_i^s, W_s$  为可训练参数,  $\sigma$  表示 Sigmoid 函数,  $y_i^s$  表示第  $i$  个单词的语义槽标签的概率.

使用 Softmax 函数获取语义槽任务的结果, 具体公式如下:

$$o_i^s = \text{Softmax}(y_i^s) \quad (7)$$

## 3.2 图交互层 (graph interaction layer)

### (1) 图的构建

根据初始层中单词级多意图识别与语义槽填充中的预测结果进行单词-意图-语义槽交互图的构建. 由于神经网络的建模结构信息能力广泛应用于各种 NLP 任务<sup>[5,15,30]</sup>. 特别是, 本文采用图注意力网络<sup>[31]</sup>构建初始层预测的单词级多意图与语义槽之间的联系, 将两个任务之间的关联性应用于最终的多意图与语义槽结果预测中. 具体交互图模型如图 3 所示, 图的构建方式如下.

• 顶点: 为了模拟意图识别与语义槽填充两个任务之间的联系, 本文定义  $V$  作为顶点的集合, 其中应用了 3 种顶点, 分别是每个单词  $H$ , 每个语义槽标签  $S$ , 每个意图标签  $I$ , 则顶点集合表示为:

$$V = [H; S; I] \quad (8)$$

其中, 单词顶点的初始表示为 BERT 输出的相应的单词的隐藏状态  $\tilde{H}_H = H \in \mathbb{R}^{N \times d}$ 、语义槽标签顶点进行随机初始化表示  $\tilde{H}_S = \{\phi^{\text{emb}}(o'_1), \dots, \phi^{\text{emb}}(o'_l)\} \in \mathbb{R}^{|S| \times d}$ 、意图标签顶点进行随机初始化表示  $\tilde{H}_I = \{\Phi^{\text{emb}}(o'_1), \dots, \Phi^{\text{emb}}(o'_l)\} \in \mathbb{R}^{|I| \times d}$ , 其中  $\phi^{\text{emb}}$  和  $\Phi^{\text{emb}}$  为可训练的嵌入矩阵, 从而获取所有节点的第 1 层状态向量:

$$\tilde{H} = [\tilde{H}_H; \tilde{H}_S; \tilde{H}_I] \in \mathbb{R}^{(N+|S|+|I|) \times d} \quad (9)$$

• 边: 由于有 3 种顶点, 存在 3 种类型的边, 对应的图邻接矩阵表示为  $\mathbf{A} \in \mathbb{R}^{(N+|S|+|I|) \times (N+|S|+|I|)}$ ; (1) 意图连接: 其中语句中的第  $i$  个单词在单词级别的意图识别对应的是第  $j$  个意图标签:  $A_{ij} = 1 (j \in o^j)$ , 对应的是图 3 中的蓝色的连接线; (2) 语义槽连接: 其中语句中的第  $i$  个单词在语义槽填充层对应的  $k$  个语义槽标签:  $A_{ik} = 1 (k \in o^s)$ , 对应的是图 3 中的绿色的连接线; (3) 意图-语义槽连接: 对于同一个单词对应第  $j$  个意图标签和第  $k$  个语义槽标签:  $A_{jk} = 1 (j \in o^j, k \in o^s)$ , 对应的是图 3 中的橙色的连接线.

根据上述顶点与边的表示构建图表示, 从而通过初始结果获取单词、语义槽标签、意图标签之间的依赖关系, 然后根据图注意力网络建模在不同节点之间的关联性.

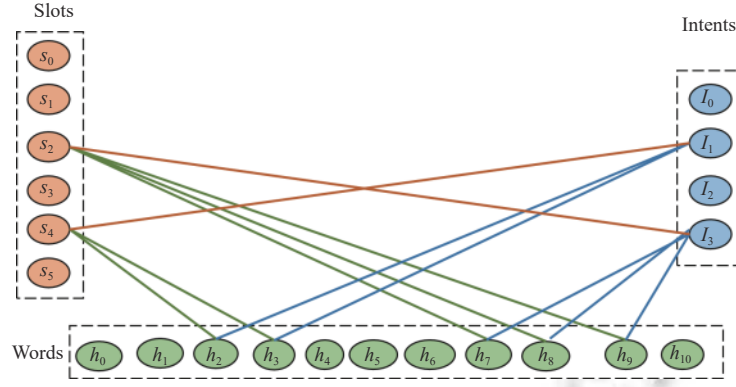


图3 图交互层

## (2) 图注意力网络 (GAT)

图注意力网络<sup>[15]</sup>是神经网络的一个变体,它融合了模型中的图结构信息和节点特征.该算法通过隐蔽的自我注意层使节点能够关注邻域特征并学习不同的注意权重,从而自动确定当前节点与其邻域之间的重要性和相关性.特别地,对于一个给定的有  $N$  个节点的图,一层 GAT 采用初始节点特征  $\tilde{H} = \{\tilde{h}_1, \dots, \tilde{h}_N\}$ ,  $\tilde{h}_n \in \mathbb{R}^F$ , 作为输入,目的是产生更抽象的表示,  $\tilde{H}' = \{\tilde{h}'_1, \dots, \tilde{h}'_N\}$ ,  $\tilde{h}'_n \in \mathbb{R}^{F'}$  作为输出. 本文对上述构建的图结构进行图注意力网络的计算,具体如公式 (10)、公式 (11) 所示:

$$\tilde{h}_i = \text{softmax}_{j \in N_i} \left( \sum_{k=1}^K \alpha_{ij}^k W_h^k \tilde{h}_j \right) \quad (10)$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [W_h \tilde{h}_i \| W_h \tilde{h}_j]))}{\sum_{j' \in N_i} \exp(\text{LeakyReLU}(a^T [W_h \tilde{h}_i \| W_h \tilde{h}_{j'}]))} \quad (11)$$

其中,  $W_h \in \mathbb{R}^{F' \times F}$  和  $a \in \mathbb{R}^{2F'}$  是可训练的权值矩阵,  $N_i$  表示节点  $i$  (包括  $i$ ) 的邻居,  $\alpha_{ij}$  是归一化注意系数,  $\sigma$  表示非线性激活函数,  $k$  表示头的数目.

## (3) 信息聚合

第  $l$  层的聚合过程可以定义为下式所示:

$$s_i^{(l+1)} = \sigma \left( \sum_{j' \in N_i} \alpha_{ij'} W_i s_{j'}^l \right) \quad (12)$$

其中,  $N_i$  表示的是一组顶点,在堆叠  $L$  层之后,我们获得上下文感知的隐藏特征  $S^{L+1} = \{s_1^{L+1}, s_2^{L+1}, \dots, s_n^{L+1}\}$ .

## 3.3 预测层

本文采用不同的单向 LSTM 作为意图识别与语义槽填充两个任务的解码器,对于意图解码器,首先从图交互层输出的结果输入到单向 LSTM,如公式 (13):

$$h_i^l = \text{LSTM}(s_i^{(l+1)}, h_{(i-1)}^l, h_{(i+1)}^l) \quad (13)$$

$$y_i^l = \text{Softmax}(h_i^l) \quad (14)$$

根据意图标签中的概率分布进行投票获取句子中的意图标签,如公式 (15)、公式 (16):

$$I_i = \begin{cases} 1, & \text{if } I_{(i,k)} > 0.5 \\ 0, & \text{others} \end{cases} \quad (15)$$

$$o^l = o_k^l \left( \left| \sum_{i=1}^n I_i \right| > \frac{n}{2} \right) \quad (16)$$

将句子级的意图与单词级的意图进行并集操作,合并得到模型预测后的最终意图结果,如公式 (17):



$$o'' = o' \cup o^I \quad (17)$$

语义槽填充解码器的公式如下所示:

$$h_i^{S'} = LSTM(s_i^{(d+1)}, h_{(i-1)}^S, h_{(i+1)}^S) \quad (18)$$

然后, 利用 Softmax 函数实现归一化, 如公式 (19), 使候选集中候选语义槽标签的总概率和为 1,  $y_i^{S'}$  为单词对应语义槽标签的概率分布.  $o^{S'}$  为概率最大的候选语义槽标签, 如公式 (20):

$$y_i^{S'} = \text{Softmax}(h_i^{S'}) \quad (19)$$

$$o^{S'} = \text{argmax}(y_i^{S'}) \quad (20)$$

### 3.4 联合训练

本文采用联合训练模型来考虑这两个任务, 并通过联合优化来更新参数. 本文对句子级意图识别、单词级意图识别、语义槽填充 3 个任务进行交叉熵损失函数计算.

首先, 计算句子级别意图识别的目标, 其具体公式如下所示:

$$L_1 = - \sum_{k=1}^{|I|} \hat{y}_k^I \log(y_k^I) + (1 - \hat{y}_k^I) \log(1 - y_k^I) \quad (21)$$

其中,  $\hat{y}_k^I$  表示用户语句准确的意图标签,  $y_k^I$  表示用户语句预测的意图标签.

其次, 计算单词级别意图识别的目标, 其具体公式如下所示:

$$CE(\hat{y}, y) = \hat{y} \log(y) + (1 - \hat{y}) \log(1 - y) \quad (22)$$

$$L_2 = - \sum_{i=1}^n \sum_{j=1}^{|I|} CE(\hat{y}_i^{(j,I)}, y_i^{(j,I)}) \quad (23)$$

其中,  $\hat{y}_i^{(j,I)}$  表示单词  $i$  准确意图标签,  $y_i^{(j,I)}$  表示单词  $i$  预测的意图标签.

最后, 计算语义槽填充任务的目标, 其具体公式如下所示:

$$L_3 = - \sum_{i=1}^n \sum_{j=1}^{|S|} \hat{y}_i^{(j,S)} \log y_i^{(j,S)} \quad (24)$$

其中,  $\hat{y}_i^{(j,S)}$  表示单词  $i$  准确语义槽标签,  $y_i^{(j,S)}$  表示单词  $i$  预测的语义槽标签,  $n$  表示语句中单词的个数,  $|I|$  表示单个意图标签的个数,  $|S|$  表示语义槽标签的个数.

学习多任务学习方式, 将多个损失函数以线性函数的方式进行组合, 最终联合训练的目标如下:

$$L = L_1 + \alpha L_2 + \beta L_3 \quad (25)$$

其中,  $\alpha$ 、 $\beta$  为超参数.

## 4 实验与结果

### 4.1 实验设置

#### • 数据集

本文在两个公开可用的多意图数据集上进行实验. 一个是 MixATIS<sup>[17,31]</sup>, 主要内容包含的是航班工作人员的录音; 另一个是 MixSNIPS<sup>[17,32]</sup>, 由个人语音助手 Snips 收集而来. 表 2 给出两个数据集对应的详细信息.

#### • 评估指标

本文采用语义槽 F1 值 (Slot F1) 作为语义槽填充任务评估指标, 意图准确性 (intent ACC) 作为多意图识别任务的评价指标, 语义准确性 (SeFr ACC) 作为句子级语义框架解析的评估指标, 其中语义准确性衡量的是一个句子中意图和语义槽均被正确语句的句子所占的比例.

#### • 实验参数

本文的预训练语言模型选择 English uncased BERT-Base 模型, 其中层数为 12、隐藏层维度为 728、头数为

12. 由于两个数据集之间在数据量、所处领域等多个方面有所不同, 因而对于两个不同的数据集, 采用不同的超参数, 具体的超参数信息如表 3 所示.

表 2 数据集信息

数据集	MixATIS	MixSNIPS
训练集	13 162	39 776
验证集	756	2 198
测试集	828	2 199
意图数量	20	7
语义槽数量	120	72
每个样本平均意图个数	2	2
单意图: 双意图: 三意图	1:6:3	2:6:2

表 3 超参数信息

超参数	MixSNIPS	MixATIS
n_heads	4	2
dropout_rate	0.6	0.1
gat_dropout_rate	0.4	0.4
n_layers_decoder	3	5
decoder_gat_hidden_dim	64	8
slot_loss_coef( $\beta$ )	2	2
Intent_token_coef( $\alpha$ )	1	1

## 4.2 基准模型

针对不同的意图场景进行模型的性能比较, 即本文模型分别和单意图识别与语义槽填充联合模型、多意图识别与语义槽填充联合模型进行比较. 在评估多意图识别任务在单意图识别与语义槽填充联合模型中的效果时, 遵循 Gangadharaiah 等人<sup>[15]</sup>、Qin 等人<sup>[16]</sup>的方法将多个意图用“#”连接成为单个意图. 基线模型的具体描述有如下 9 种.

- (1) Attention BiRNN: Liu 等人<sup>[20]</sup>提出了一种基于对齐的 RNN, 用于联合语义槽填充和意图识别.
  - (2) Slot-gated: Goo 等人<sup>[5]</sup>提出了一个语义槽门控联合模型, 明确考虑了语义槽填充和意图识别之间的相关性.
  - (3) Bi-Model: Wang 等人<sup>[2]</sup>提出了双向模型来模拟意图识别和语义槽填充之间的双向关系.
  - (4) SF-ID: E 等人<sup>[7]</sup>提出了 SF-ID 网络来建立两个任务之间的直接连接.
  - (5) Stack-propagation: Qin 等人<sup>[23]</sup>采用了一种栈传播框架, 明确地将意图识别引导槽填充, 从而实现二者任务的联系.
  - (6) Joint Multiple ID-SF: Gangadharaiah 等人<sup>[15]</sup>提出了一个带有语义槽门机制的多任务框架, 用于多意图识别和语义槽填充.
  - (7) AGIF: Qin 等人<sup>[16]</sup>提出了一个自适应交互网络, 以实现细粒度的多意图信息集成.
  - (8) GL-GIN: Qin 等人<sup>[17]</sup>提出一种全局-局部交互网络, 尝试探索联合多意图识别和槽填充的非自回归方法.
  - (9) SLIM: Cai 等人<sup>[33]</sup>提出基于 BERT 联合学习多意图识别和语义槽填充.
- 在 MixATIS 和 MixSNIPS 上, 我们比较了单意图模型: (1)–(5), 同时比较了多意图模型: (6)–(9).

## 4.3 实验结果

### (1) 实验结果对比分析

表 4 展示了不同模型在 MixSNIPS 和 MixATIS 两个数据集上的性能. 从表 4 的结果上, 我们可以看出: 1) 本文模型在两个数据集上相对于基线模型均有一定的提高, 这证明了本文模型的有效性; 2) 与 SLIM 模型相比, 在 MixSNIPS 数据集上, Slot F1 提高了 0.72%, Intent ACC 提高了 0.86%, 在 MixATIS 数据集上 Intent ACC (+4.42%), Slot F1 指标未达到最好的原因是 MixATIS 数据集的全部意图都是关于航班信息, 各种词汇之间存在着一定的相似性, 当图交互网络进行信息传递时, 意图之间的语义信息可能混淆性, 同时该数据集中意图之间的样本数据存在高度不均衡性; 3) 在语义准确性上, 相比于 SLIM 模型, MixSNIPS 数据集上提高了 2.58%, MixATIS 数据集上提高了 3.53%, 表明单词-意图-语义槽交互图中的一对一对应关系和建立的意图与语义槽之间的联系性为整体语义准确性提供了保证.

### (2) 消融实验分析

表 5 主要通过消融实验对模型中句子级别意图识别、单词级别意图识别、单词-意图-语义槽交互层这 3 个组件进行分析, 其中 w/o 表示完整模型去除该组件后的实验结果.

表4 对比实验分析 (%)

Model	MixSNIPS			MixATIS		
	Slot F1	Intent ACC	SeFr ACC	Slot F1	Intent ACC	SeFr ACC
Attention BiRNN* (2016)	89.4	94.1	62.2	86.4	74.6	39.1
Slot-gated* (2018)	87.8	96.0	56.5	87.7	63.9	35.5
Bi-model* (2018)	86.8	95.3	53.9	83.9	70.3	34.4
SF-ID* (2019)	89.6	96.3	59.3	87.4	66.2	34.9
Stack-propagation* (2019)	93.9	96.4	75.5	87.8	72.1	40.1
Joint Multiple ID-SF <sup>†</sup> (2019)	91.0	95.7	66.6	84.6	73.4	36.1
AGIF <sup>†</sup> (2020)	94.5	96.5	76.4	86.7	74.4	40.8
GL-GIN <sup>†</sup> (2021)	94.9	95.6	75.4	88.3	76.3	43.5
SLIM <sup>†</sup> (2022)	96.5	97.2	84.0	<b>88.5</b>	78.3	47.6
WISM	<b>97.2</b>	<b>98.0</b>	<b>86.2</b>	88.1	<b>81.8</b>	<b>49.3</b>

注: \*表示为单意图模型; †表示为多意图模型; 表格中的结果采用四舍五入的方式保留1位有效数字

表5 消融实验分析 (%)

Model	MixSNIPS			MixATIS		
	Slot F1	Intent ACC	SeFr ACC	Slot F1	Intent ACC	SeFr ACC
w/o intent classifier	96.9	96.9	85.0	86.5	81.2	48.3
w/o intent token classifier	96.7	96.8	85.1	86.2	79.9	48.2
w/o graph interaction layer	96.9	96.6	84.8	86.1	78.1	45.3
w/o 投票机制	97.1	97.3	85.4	87.5	80.0	48.7
WISM	97.2	98.0	86.2	88.1	81.8	49.3

通过表5可以看出, (1) 当去除句子级意图识别时, MixSNIPS 数据集中的 Slot F1、Intent ACC、SeFr ACC 这3个指标分别下降了0.27%、1.2%、1.3%, 在 MixATIS 数据集中3个指标分别下降了1.82%、0.74%、1.98%; (2) 当去除单词级意图识别时, 可以观察到在 MixSNIPS 和 MixATIS 数据集中 Slot F1 分别降低了0.50%和2.13%, Intent ACC 降低了1.29%和2.33%, SeFr ACC 降低了1.26%和2.23%; (3) 当图交互层时, 在 MixSNIPS 数据集中 Slot F1 (-0.31%)、Intent ACC (-1.43%)、SeFr ACC (-1.57%), 在 MixATIS 数据集中 Slot F1 (-2.29%)、Intent ACC (-4.42%)、SeFr ACC (-8.09%), (4) w/o 投票机制表示去除单词级意图识别模块中的投票机制, 考虑投票机制对于最终模型性能的影响, 在 MixSNIPS 数据集3个指标分别下降了0.10%、0.71%、0.93%, 在 MixATIS 数据集中的 Slot F1、Intent ACC、SeFr ACC 这3个指标分别下降了0.68%、2.20%、1.22%。

通过对消融实验结果的分析可以得出, 在句子级意图识别、单词级意图识别、图交互层3个组件中, 图交互层对句子语义准确性的影响最为明显. 句子级意图识别与单词级意图识别两个组件的删除对于多意图识别任务均存在着性能的降低, 可以从另一角度呈现了单词级与句子级两个维度的意图识别任务同时为整体语义准确性提供了保证; 并且额外考虑单词级意图识别模块中投票机制对于模型性能的影响, 可以通过表5看出投票机制虽然只是模型中较为简单的模块, 其对意图识别与语义槽填充两个任务均有着一定的影响, 为模型的整体性能提供了保障。

### (3) 单词-意图-识别交互层分析

表6对图注意力网络中的顶点、边进行分析, 从而进一步分析两个任务之间的联系性. 通过表6可以看出, (1) 当去除意图连接时, 即只考虑语义槽对于意图的单向联系性, 此时在 MixSNIPS 数据集中 Slot F1 (-0.15%), Intent ACC (-0.65%), SeFr ACC (-0.42%), 在 MixATIS 数据集中, Slot F1 降低了2.65%, Intent ACC 降低了1.64%, SeFr ACC 降低了6.63%; (2) 当去除语义槽连接时, 即仅仅考虑意图对语义槽的单向联系性, 此时在 MixSNIPS 数据集中 Slot F1 降低了0.01%, Intent ACC 降低了1.14%, SeFr ACC 降低了1.86%, 在 MixATIS 数据集中 Slot F1 (-1.86%), Intent ACC (-2.69%), SeFr ACC (-5.64%); (3) 当去除意图与语义槽的连接时, 即只保留意图和语义槽分

别对单词的连接,此时在 MixSNIPS 数据集中的 3 个指标分别下降了 0.73%、0.92%、2.47%,在 MixATIS 数据集中的 3 个指标分别下降了 1.98%、7.08%、7.85%;(4) 当去除意图节点时,即完全去除意图标签的作用,可以观察到在 MixSNIPS 数据集和 MixATIS 数据集中 3 个指标的下降情况分别是: Slot F1 (-0.57%、-2.99%), Intent ACC (-1.44%、-5.61%), SeFr ACC (-1.90%、-9.07%);(5) 当去除语义槽节点时,即完全去除语义槽标签的作用,可以观察到在 MixSNIPS 和 MixATIS 数据集中 Slot F1 分别降低了 0.76% 和 2.62%, Intent ACC 降低了 1.57% 和 6.34%, SeFr ACC 降低了 3.26% 和 6.14%.

表 6 意图-语义槽-单词图网络分析 (%)

Model	MixSNIPS			MixATIS		
	Slot F1	Intent ACC	SeFr ACC	Slot F1	Intent ACC	SeFr ACC
w/o intent edge	96.8	97.0	84.7	85.1	79.5	46.0
w/o slot edge	96.9	96.2	84.2	85.7	78.6	46.5
w/o intent slot edge	96.5	97.1	84.0	86.4	76.0	45.4
w/o intent node	96.6	96.6	84.5	85.5	77.2	44.8
w/o slot node	96.5	96.5	83.4	85.8	76.6	46.3
WISM	97.2	98.0	86.2	88.1	81.8	49.3

通过上述情况进行分析,可以了解到当去除意图连接时,表明去除了该意图对该对应的语义槽的影响,从而此时对语义槽的影响较大,当去除语义槽连接时此时消除语义槽任务对于意图任务的影响,结果表明主要影响意图的结果,但是无论去除某一部分的连接对语义准确性均有一定的影响,表示两个任务之间的联系性的利用提升了模型的性能;当去除某一节点时,此时对于完全的去其标签对于结果的影响,可以观察到当去除某一节点时,对最终结果的影响更为显著,因为去除该节点的同时消除了任务之间的完全联系性;该消融实验体现了两个任务之间是彼此影响的,同时说明二者之间的联系性对任务结果有着一定的积极意义.

#### (4) 参数影响分析

为了充分探讨算法受模型参数的影响情况,分别针对图注意力网络中和意图识别模块中的不同参数权重进行实验,用来分析其参数对模型性能所带来的影响.

##### 1) 注意力头数影响

本文算法中的图注意力网络中采用了多头注意力丰富模型能力并稳定学习过程,本部分在 MixSNIPS 和 MixATIS 两个数据集上对图注意力网络中注意力头数这一超参数对语义准确性进行研究,具体结果如图 4 所示.可以观察到在 MixSNIPS 数据集中当注意力头数为 4 时语义准确性最高,从而在本文中对于 MixSNIPS 数据集选择注意力头数为 4;在 MixATIS 数据集中当注意力头数为 2 时语义准确性最高,从而在本文中对于 MixATIS 数据集选择注意力头数为 2.

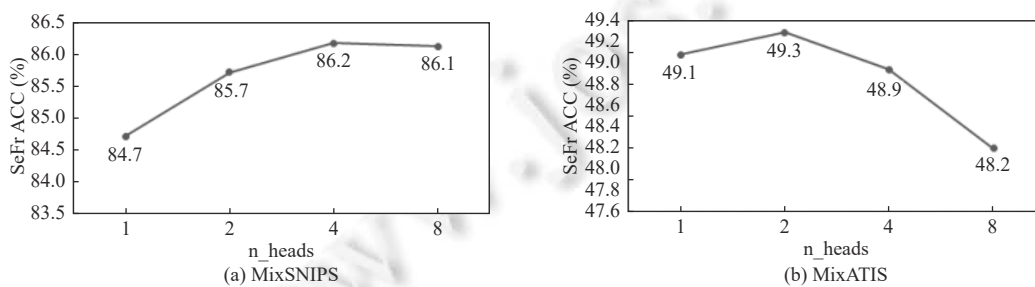


图 4 图注意力头数影响

##### 2) 意图识别模块分析

本文算法中的意图识别模块采用句子级意图识别与单词级意图识别结合的方式得到最终的意图识别结果.该

部分首先在 MixSNIPS 和 MixATIS 两个数据集上对句子级意图识别中的阈值  $\delta$  这一超参数对语义准确性进行研究, 具体结果如图 5 所示. 可以观察到在 MixSNIPS 和 MixATIS 两个数据集中当阈值  $\delta$  为 0.5 时句子语义准确性最高, 从而本文在句子级意图识别模型中阈值选择为 0.5.

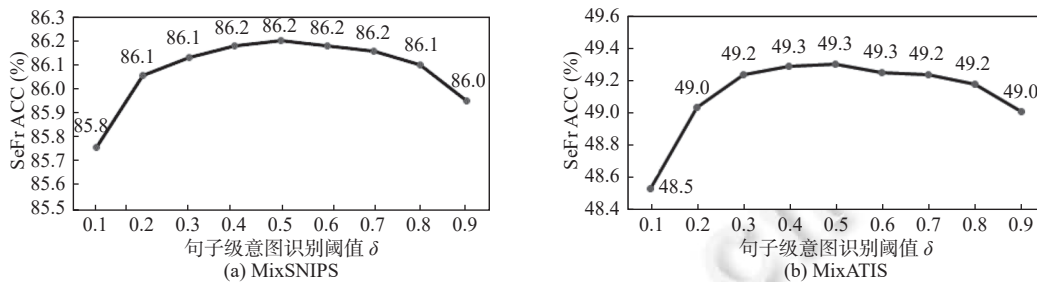


图 5 句子级意图识别阈值  $\delta$  影响

在单词级意图识别模块中, 对于单词级的意图结果本文算法中使用的是投票机制获取整个句子的意图, 该部分实验单词级意图识别模块中投票肯定票阈值的设置对最终模型语义准确性的研究, 模型在 MixSNIPS 和 MixATIS 两个数据集均进行实验, 具体的实验结果如图 6 所示. 在 MixSNIPS 和 MixATIS 两个数据集上均在投票肯定票阈值为 0.5 时模型的语义准确性最高, 因而本文模型在单词级意图识别模块对于肯定票的阈值设置为 0.5.

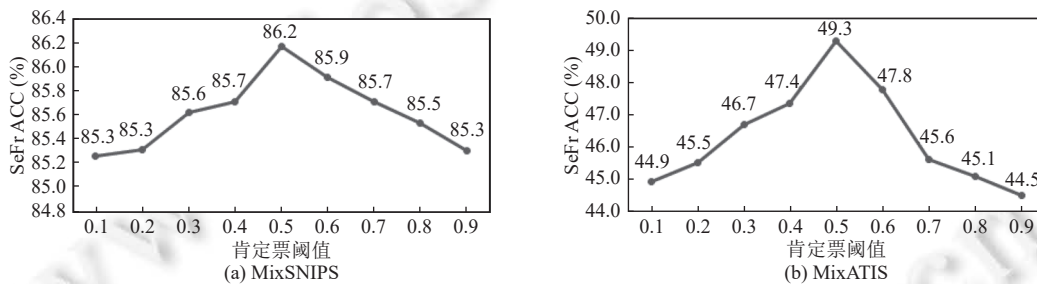


图 6 单词级意图识别肯定票阈值影响

##### (5) 细粒度意图影响

选择 MixATIS 数据集中测试集的一句话展现在本论文中细粒度意图分别对于语义槽和句子中的单词的影响, 选择的句子是: show me the cheapest round trip fares from san francisco to houston and then how many passengers can an 11011 aircraft hold, 其对应的语义槽标签是: O O O B-cost\_relative B-round\_trip I-round\_trip O O B-fromloc.city\_name I-fromloc.city\_name O B-toloc.city\_name O O O O O O B-aircraft\_code O O, 其意图标签是: atis\_airfare 和 atis\_capacity.

图 7 展示了细粒度意图标签对于语义槽标签的影响. 可以观察到在该语句中, atis\_airfare 意图标签主要对 B-aircraft\_code 语义槽标签有着一定的影响力, atis\_capacity 意图标签主要对 B-cost\_relative、B-round\_trip、I-round\_trip、B-fromloc.city\_name、I-fromloc.city\_name、B-toloc.city\_name 这 6 个语义槽标签有着一定的影响力.

图 8 展示了细粒度意图对于句子中的单词的影响. 在图 8 中显示了句子中的两个意图标签分别对句子中的单词的影响力, 可以在图 8 中观察到 atis\_airfare 意图标签主要对 11011 这一单词具有一定的影响力; atis\_capacity 意图标签分别对于 cheapest、round、trip、fares、san、francisco、houston 这 6 个单词具有一定的影响力.

将图 7 与图 8 两个意图标签具有影响力的语义槽与单词进行联系, 可以看出与意图标签对应有影响力的语句单词可以在同一意图标签对应有影响力的语义槽标签中找到与之匹配的语义槽标签, 可以了解到意图标签、语义槽标签、语句单词三者之间的联系性, 从而佐证其细粒度单词级意图与语义槽任务联系的重要性.

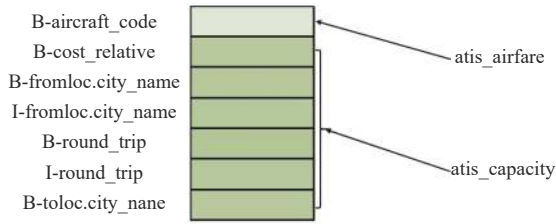


图 7 细粒度意图对语义槽影响

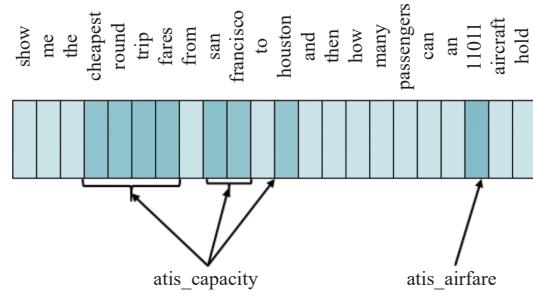


图 8 细粒度意图对单词影响

(6) 语义槽隐藏簇状态可视化

为了展示语义槽填充任务在本模型中的有效性和更好地理解本模型的最终的可靠的结果, 我们对 MixSNIPS 测试集中的话语词的最终隐藏状态进行了 TSNE<sup>[34]</sup> 可视化. 因为单个隐藏状态通常对应于多个意图, 很难将意图状态可视化; 因此, 我们可视化了话语中每个单词的最终语义槽隐藏状态. 为了简化可视化, 我们随机选择 7 个语义槽标签 (剔除 O 标签) 进行显示, 其中包含有 2 500 多个单词, 其语义槽标签的簇如图 9 所示.

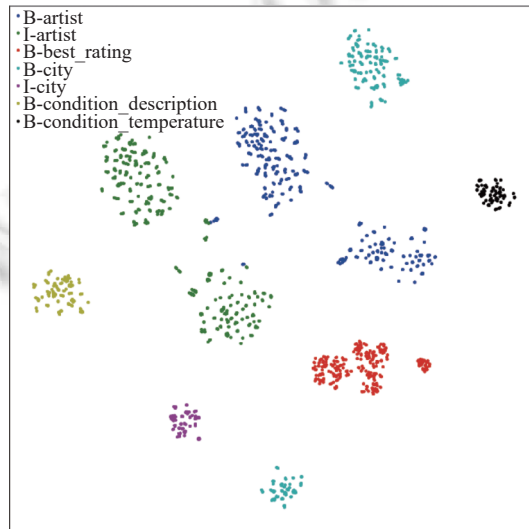


图 9 语义槽状态可视化

从图 9 中, 我们可以观察到聚类的边界是清晰的, 同一聚类中的数据点之间是紧密封闭的, 我们还可以发现, B 槽集群和它们对应的 I 槽集群是明显分开的.

上述观察结果证明, 本文方法可以通过有效地捕获和利用意图与语义槽标签相关性来更好地学习隐藏状态.

5 总 结

对话系统旨在解析用户输入语句并做出相应的回答, 因而能够准确理解用户的语义表达显得尤为重要. 多意图识别与语义槽填充联合模型在将用户输入语句转化为机器理解的语义表达展现了不错的效果. 但现有联合模型在意图识别与语义槽填充两个任务交互过程中存在错误传播问题和错误映射关系. 为了解决上述问题, 本文提出了基于图注意力网络的多意图识别与语义槽填充联合模型 (WISM), 模型首先采用单词级意图与语义槽形成一对一映射关系明确二者之间关系以解决意图信息与语义槽信息之间的错误对应关系, 然后构建单词-意图-语义槽交互图, 随后通过图注意力网络建立两个任务之间单词级别的双向联系性, 从而降低交互过程中错误传播带来的影

响。WISM 模型在 MixSNIPS 和 MixATIS 数据集上相较于 SLIM 语义准确性分别提高了 2.58% 和 3.53%, 并通过消融实验和一系列可视化实验验证了模型的有效性。未来, 我们将进一步深入研究如何将标签语义信息融入模型中<sup>[35,36]</sup>。

## References:

- [1] Hou LX, Li YL, Li CC. Review of research on task-oriented spoken language understanding. *Computer Engineering and Applications*, 2019, 55(11): 7–15 (in Chinese with English abstract). [doi: [10.3778/j.issn.1002-8331.1901-0056](https://doi.org/10.3778/j.issn.1002-8331.1901-0056)]
- [2] Wang Y, Shen YL, Jin HX. A bi-model based RNN semantic frame parsing model for intent detection and slot filling. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: Association for Computational Linguistics, 2018. 309–314. [doi: [10.18653/v1/N18-2050](https://doi.org/10.18653/v1/N18-2050)]
- [3] Zhang XD, Wang HF. A joint model of intent determination and slot filling for spoken language understanding. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence. New York: AAAI Press, 2016. 2993–2999.
- [4] Wei PF, Zeng B, Wang MH, Zeng A. Survey on joint modeling algorithms for spoken language understanding based on deep learning. *Ruan Jian Xue Bao/Journal of Software*, 2022, 33(11): 4192–4216 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6385.htm> [doi: [10.13328/j.cnki.jos.006385](https://doi.org/10.13328/j.cnki.jos.006385)]
- [5] Goo CW, Gao G, Hsu YK, Huo CL, Chen TC, Hsu KW, Chen YN. Slot-gated modeling for joint slot filling and intent prediction. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: Association for Computational Linguistics, 2018. 753–757. [doi: [10.18653/v1/N18-2118](https://doi.org/10.18653/v1/N18-2118)]
- [6] Li CL, Li L, Qi J. A self-attentive model with gate mechanism for spoken language understanding. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 3824–3833. [doi: [10.18653/v1/D18-1417](https://doi.org/10.18653/v1/D18-1417)]
- [7] E HH, Niu PQ, Chen ZF, Song MN. A novel bi-directional interrelated model for joint intent detection and slot filling. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 5467–5471. [doi: [10.18653/v1/P19-1544](https://doi.org/10.18653/v1/P19-1544)]
- [8] Zhang CW, Li YL, Du N, Fan W, Yu PS. Joint slot filling and intent detection via capsule neural networks. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 5259–5267. [doi: [10.18653/v1/P19-1519](https://doi.org/10.18653/v1/P19-1519)]
- [9] Liu YJ, Meng FD, Zhang JC, Zhou J, Chen YF, Xu JN. CM-Net: A novel collaborative memory network for spoken language understanding. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 1051–1060. [doi: [10.18653/v1/D19-1097](https://doi.org/10.18653/v1/D19-1097)]
- [10] Tang H, Ji DH, Zhou QJ. End-to-end masked graph-based CRF for joint slot filling and intent detection. *Neurocomputing*, 2020, 413: 348–359. [doi: [10.1016/j.neucom.2020.06.113](https://doi.org/10.1016/j.neucom.2020.06.113)]
- [11] Wei PF, Zeng B, Liao WX. Joint intent detection and slot filling with wheel-graph attention networks. *Journal of Intelligent & Fuzzy Systems*, 2022, 42(3): 2409–2420. [doi: [10.3233/JIFS-211674](https://doi.org/10.3233/JIFS-211674)]
- [12] Liu F, Perez J. Gated end-to-end memory networks. In: Proc. of the 15th Conf. of the European Chapter of the Association for Computational Linguistics. Valencia: Association for Computational Linguistics, 2017. 1–10. [doi: [10.18653/v1/e17-1001](https://doi.org/10.18653/v1/e17-1001)]
- [13] Kim B, Ryu S, Lee GG. Two-stage multi-intent detection for spoken language understanding. *Multimedia Tools and Applications*, 2017, 76(9): 11377–11390. [doi: [10.1007/s11042-016-3724-4](https://doi.org/10.1007/s11042-016-3724-4)]
- [14] Xu PY, Sarikaya R. Exploiting shared information for multi-intent natural language sentence classification. In: Proc. of the 14th Annual Conf. of the Int'l Speech Communication Association. Lyon, 2013. 3785–3789. [doi: [10.21437/Interspeech.2013-599](https://doi.org/10.21437/Interspeech.2013-599)]
- [15] Gangadharaiah R, Narayanaswamy B. Joint multiple intent detection and slot labeling for goal-oriented dialog. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019. 564–569. [doi: [10.18653/v1/n19-1055](https://doi.org/10.18653/v1/n19-1055)]
- [16] Qin LB, Xu X, Che WX, Liu T. AGIF: An adaptive graph-interactive framework for joint multiple intent detection and slot filling. In: Proc. of the 2020 Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics, 2020. 1807–1816. [doi: [10.18653/v1/2020.findings-emnlp.163](https://doi.org/10.18653/v1/2020.findings-emnlp.163)]
- [17] Qin LB, Wei FX, Xie TB, Xu X, Che WX, Liu T. GL-GIN: Fast and accurate non-autoregressive model for joint multiple intent detection and slot filling. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on

- Natural Language Processing. Association for Computational Linguistics, 2021. 178–188. [doi: [10.18653/v1/2021.acl-long.15](https://doi.org/10.18653/v1/2021.acl-long.15)]
- [18] Guo D, Tur G, Yih WT, Zweig G. Joint semantic utterance classification and slot filling with recursive neural networks. In: Proc. of the 2014 IEEE Spoken Language Technology Workshop (SLT). South Lake Tahoe: IEEE, 2014. 554–559. [doi: [10.1109/SLT.2014.7078634](https://doi.org/10.1109/SLT.2014.7078634)]
- [19] Zhou QR, Wen LY, Wang XJ, Ma L, Wang Y. A hierarchical LSTM model for joint tasks. In: Proc. of the 15th China National Conf. and 4th Int'l Symp. on Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Yantai: Springer, 2016. 324–335. [doi: [10.1007/978-3-319-47674-2\\_27](https://doi.org/10.1007/978-3-319-47674-2_27)]
- [20] Liu B, Lane I. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. In: Proc. of the 17th Annual Conf. of the Int'l Speech Communication Association (Interspeech 2016). 2016. 685–689. [doi: [10.21437/Interspeech.2016-1352](https://doi.org/10.21437/Interspeech.2016-1352)]
- [21] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- [22] Chen MY, Zeng J, Lou J. A self-attention joint model for spoken language understanding in situational dialog applications. arXiv:1905.11393, 2019.
- [23] Qin LB, Che WX, Li YM, Wen HY, Liu T. A stack-propagation framework with token-level intent detection for spoken language understanding. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019. 2078–2087. [doi: [10.18653/v1/D19-1214](https://doi.org/10.18653/v1/D19-1214)]
- [24] Wang CR, Huang Z, Hu MH. SASGBC: Improving sequence labeling performance for joint learning of slot filling and intent detection. In: Proc. of the 6th Int'l Conf. on Computing and Data Engineering. Sanya: ACM, 2020. 29–33. [doi: [10.1145/3379247.3379266](https://doi.org/10.1145/3379247.3379266)]
- [25] Han SC, Long SQ, Li HC, Weld H, Poon J. Bi-directional joint neural networks for intent classification and slot filling. In: Proc. of the 22nd Annual Conf. of the Int'l Speech Communication Association (InterSpeech 2021). Brno: ISCA, 2021. 4743–4747. [doi: [10.21437/Interspeech.2021-2044](https://doi.org/10.21437/Interspeech.2021-2044)]
- [26] Zhang LH, Ma DH, Zhang XD, Yan XH, Wang HF. Graph LSTM with context-gated mechanism for spoken language understanding. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conf., the 10th AAAI Symp. on Educational Advances in Artificial Intelligence. New York: AAAI Press, 2020. 9539–9546. [doi: [10.1609/aaai.v34i05.6499](https://doi.org/10.1609/aaai.v34i05.6499)]
- [27] Tao SM, Qin Y, Chen YM, Du CN, Sun HF, Meng WB, Xiao YH, Guo JX, Su C, Wang MH, Zhang M, Wang YX, Yang H. Incorporating complete syntactical knowledge for spoken language understanding. In: Proc. of the 6th China Conf. on Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction. Guangzhou: Springer, 2021. 145–156. [doi: [10.1007/978-981-16-6471-7\\_11](https://doi.org/10.1007/978-981-16-6471-7_11)]
- [28] Wu J, Harris IG, Zhao HZ, Ling GM. A graph-to-sequence model for joint intent detection and slot filling. In: Proc. of the 17th IEEE Int'l Conf. on Semantic Computing (ICSC). Laguna Hills: IEEE, 2023. 131–138. [doi: [10.1109/ICSC56153.2023.00028](https://doi.org/10.1109/ICSC56153.2023.00028)]
- [29] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019. 4171–4186. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
- [30] Shen T, Zhou TY, Long GD, Jiang J, Pan SR, Zhang CQ. DiSAN: Directional self-attention network for RNN/CNN-free language understanding. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symp. on Educational Advances in Artificial Intelligence. New Orleans: AAAI Press, 2018. 5446–5455. [doi: [10.1609/aaai.v32i1.11941](https://doi.org/10.1609/aaai.v32i1.11941)]
- [31] Hemphill CT, Godfrey JJ, Doddington GR. The ATIS spoken language systems pilot corpus. In: Proc. of the 1990 Workshop on Speech and Natural Language. Hidden Valley: Association for Computational Linguistics, 1990. 96–101. [doi: [10.3115/116580.116613](https://doi.org/10.3115/116580.116613)]
- [32] Coucke A, Saade A, Ball A, Bluche T, Caulier A, Leroy D, Doumouro C, Gisselbrecht T, Caltagirone F, Lavril T, Prinet M, Dureau J. Snips voice platform: An embedded spoken language understanding system for private-by-design voice interfaces. arXiv:1805.10190, 2018.
- [33] Cai FY, Zhou WH, Mi F, Faltings B. Slim: Explicit slot-intent mapping with BERT for joint multi-intent detection and slot filling. In: Proc. of the 2022 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP). Singapore: IEEE, 2022. 7607–7611. [doi: [10.1109/ICASSP43922.2022.9747477](https://doi.org/10.1109/ICASSP43922.2022.9747477)]
- [34] van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008, 9(86): 2579–2605.
- [35] Xiong YJ, Feng YK, Wu H, Kamigaito H, Okumura M. Fusing label embedding into BERT: An efficient improvement for text classification. In: Proc. of 2021 the Association for Computational Linguistics: ACL-IJCNLP. Association for Computational Linguistics, 2021. 1743–1750. [doi: [10.18653/v1/2021.findings-acl.152](https://doi.org/10.18653/v1/2021.findings-acl.152)]
- [36] Chen CY, Hung TM, Hsu YL, Ku LW. Label-aware hyperbolic embeddings for fine-grained emotion classification. In: Proc. of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto: Association for Computational Linguistics, 2023.



10947–10958. [doi: [10.18653/v1/2023.acl-long.613](https://doi.org/10.18653/v1/2023.acl-long.613)]

附中文参考文献:

- [1] 侯丽仙, 李艳玲, 李成城. 面向任务口语理解研究现状综述. 计算机工程与应用, 2019, 55(11): 7–15. [doi: [10.3778/j.issn.1002-8331.1901-0056](https://doi.org/10.3778/j.issn.1002-8331.1901-0056)]
- [4] 魏鹏飞, 曾碧, 汪明慧, 曾安. 基于深度学习的口语理解联合建模算法综述. 软件学报, 2022, 33(11): 4192–4216. <http://www.jos.org.cn/1000-9825/6385.htm> [doi: [10.13328/j.cnki.jos.006385](https://doi.org/10.13328/j.cnki.jos.006385)]



张玉慧(1999—), 女, 硕士生, CCF 学生会员, 主要研究领域为数据挖掘, 自然语言处理, 口语理解.



据生根(1970—), 男, 博士, 教授, CCF 高级会员, 主要研究领域为数据挖掘, 自然语言处理, 知识图谱.



陈黎(1977—), 女, 博士, 讲师, 主要研究领域为自然语言处理, 文本挖掘与信息抽取, 问答系统.



李每文(2000—), 男, 硕士生, CCF 学生会员, 主要研究领域为自然语言处理, 会话式问答.