

基于多意图融合框架的联合意图识别和槽填充

尹商鉴, 黄沛杰*, 梁栋柱, 何卓棋, 黎倩尔, 徐禹洪

华南农业大学, 数学与信息学院, 广东广州, 510642

s jy8460@163.com, p jhuang@scau.edu.cn, liang_dz@stu.scau.edu.cn,
13428897035@163.com, li@stu.scau.edu.cn, xuyuhong@scau.edu.cn

摘要

近年来, 多意图口语理解 (SLU) 已经成为自然语言处理领域的研究热点。当前先进的多意图SLU模型采用图-交互式框架进行联合多意图识别和槽位填充, 能够有效地捕捉到词元级槽位填充任务的细粒度意图信息, 取得了良好的性能。但是, 它忽略了联合作用下的意图所包含的丰富信息, 没有充分利用多意图信息对槽填充任务进行指引。为此, 本文提出了一种基于多意图融合框架 (MIFF) 的联合多意图识别和槽填充框架, 使得模型能够在准确地识别不同意图的同时, 利用意图信息为槽填充任务提供更充分的指引。我们在MixATIS和MixSNIPS两个公共数据集上进行了实验, 结果表明, 我们的模型在性能和效率方面均超过了当前最先进的方法, 同时能够有效从单领域数据集泛化到多领域数据集上。

关键词: 多意图口语理解; 多意图融合框架; 联合多意图识别和槽位填充

A Multi-Intent Fusion Framework for Joint Intent Detection and Slot Filling

Shangjian Yin, Peijie Huang*, Dongzhu Liang, Zhuoqi He,
Qianer Li, Yuhong Xu

College of Mathematics and Informatics, South China Agricultural University, China
s jy8460@163.com, p jhuang@scau.edu.cn, liang_dz@stu.scau.edu.cn,
13428897035@163.com, li@stu.scau.edu.cn, xuyuhong@scau.edu.cn

Abstract

In recent years, multi-intent spoken language understanding (SLU) has become a research hotspot in the field of natural language processing. The current state-of-the-art multi-intent SLU model uses a graph-interaction framework for joint multi-intent detection and slot-filling, which can effectively capture fine-grained intent information for lexical element-level slot-filling tasks and achieves good performance. However, it ignores the rich information contained in the intent under joint action and does not fully utilize the multi-intent information to guide the slot-filling task. To this end, this paper proposes a joint multi-intent detection and slot-filling approach based on the multi-intent fusion framework (MIFF), which enables the model to accurately identify different intents while using the intent information to provide more adequate guidance for the slot-filling task. We conducted experiments on two public datasets, MixATIS and MixSNIPS, and the results show that our model outperforms current state-of-the-art methods in terms of performance and efficiency, while being able to effectively generalize from single-domain dataset to multi-domain dataset.

*通讯作者

©2023 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

Keywords: Spoken language understanding , Multi-intent fusion framework , Joint intent detection and slot filling.

1 引言

口语理解 (Spoken Language Understanding, SLU) (Tur et al., 2011; Young et al., 2013) 是面向任务的对话的关键组成部分系统, 其目的是创建一个语义框架, 总结用户的请求。利用意图识别来识别用户意图, 利用槽填充来提取相关语义成分, 构建语义框架。由于意图识别和槽填充这两个子任务紧密相关, 主流SLU系统采用联合模型来建模它们之间的相关性。在现实场景中, 用户通常在话语中表达多个意图, 如亚马逊内部数据集中52%的例子是多意图的(Gangadharaiah et al., 2019)。Figure 1展示了一个两意图示例, 其中包含一个分类任务来对意图标签进行分类(即, 预测意图为`atis_aircraft` 和`atis_city`) 和一个序列标记任务来预测槽标签序列(即, 将语句标记为 $\{0, 0, 0, 0, 0, 0, 0, 0, \text{B-aircraft_code}, 0, 0, 0, 0, 0, 0, \text{B-city_name}, 0\}$)。然而, 以往的工作大多只关注简单的单意图场景, 无法有效处理原始网络的多意图。最近, 多意图SLU逐渐受到关注, 因为它在我们日常生活中具有重要的应用价值 (Gangadharaiah et al., 2019; Qin et al., 2000), 它可以处理包含多个意图的语句。为了满足现实生活的需要, Xu等人 (2013)和Kim等人 (2017)开始探索多意图SLU。然而, 他们的模型只考虑了多重意图识别, 而忽略了槽填充任务。最近, Gangadharaiah 和Narayanaswamy (2019) 首次尝试提出了一个多任务框架来联合建模多重意图识别和槽填充。Qin等 (2020)进一步提出了一种自适应交互框架 (AGIF) 来实现细粒度的多意图信息集成。基于图注意力网络[7]和非自回归方式的思想, Qin等人 (2021)提出了一个全局局部图交互网络(global - local graph Interaction network, GL-GIN), 该网络建模了多个意图和话语中所有槽之间的槽依赖和交互, 获得了先进的性能。虽然该模型取得了较好的效果, 然而, 它在模型浅层交互中只是将多意图和槽填充信息简单的拼接, 没有充分挖掘意图中对槽的填充具有重要的指导意义的信息, 从而让训练效率和模型性能大打折扣。

为了进一步增强多意图与槽填充的关联性以及协同优化的效果, 我们提出了一个多意图融合框架, 其核心是多意图融合层, 该层包含了三种信息的融合过程, 实现了在意图识别和槽填充任务之间建立更加稳定有效的连接。考虑到意图识别任务在联合效应中起着重要作用, 我们提出了意图强化层, 获得了更丰富的多意图信息表示, 提高了整体框架的鲁棒性。在MixATIS (Hemphill et al., 1990)和MixSNIPS (Coucke et al., 2018)两个公共数据集上的实验结果表明, 我们的框架获得了最先进的性能。此外, 由于充分利用了意图信息, 我们的模型更容易协调意图识别和槽填充任务, 大大提高了训练效率。

我们的贡献总结如下: 1) 我们构建了一个意图强化层, 更好地丰富了多意图信息和缓解信息遗忘问题。2) 为了更好地利用意图信息, 我们首次(在我们认知下)尝试构建了用于联合多意图识别和槽填充的多意图融合层。3) 在两个公开数据集上的实验结果表明, 我们的框架不仅达到了最先进的性能, 而且更容易协调意图识别和槽填充, 大大提高了训练效率, 为对话系统的下游任务提供了一个创新的方法。

2 相关工作

尽管SLU已经有很长的历史, 但对于多意图SLU的研究在近几年才出现作为一个新任务。多意图SLU是由Gangadharaiah等人 (2019) 首先提出的, 并由Qin等人(2020; 2021)证明了其重要性。在这一节中, 我们将介绍SLU历史中一些经典的工作。SLU任务通常由两个子任务组成, 分别是槽填充和意图识别。槽填充任务可以被视为一个序列标注任务, 而意图识别任务可以被视为一个分类任务。传统的方法包括用条件随机场 (Conditional Random Fields, CRF) (Raymond et al., 2007)用于槽填充, 以及用于意图识别的支持向量机 (Support Vector Machines, SVM) (Haffner et al., 2003)和Adaboost (Schapire et al., 2000), 都取得了很好的效果。自从深度学习开始蓬勃发展以来, 这两个子任务的性能也达到了一个更高的水平。循环神经网络 (Recurrent Neural Network, RNN) 首次被Yao等人引入SLU任务 (Yao et al., 2000)。随后, Yao等人(2014)在SLU任务中利用了长短期记忆网络 (Long Short Term Memory, LSTM) 的优势。他们都取得了显著的效果。从此, SLU任务步入了深度学习的时

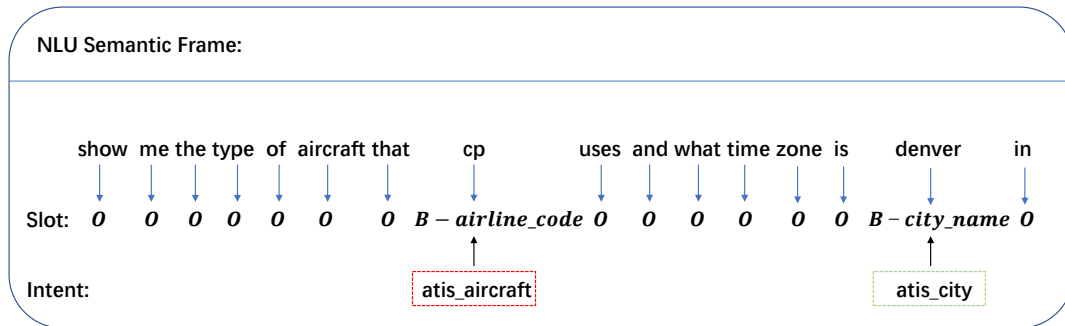


Figure 1: 多意图SLU框架

代。近年来，多任务训练，如联合学习考虑了意图与槽之间的相互关联，取得了很好的效果，推动了SLU的发展。Zhang等人(2016)通过引入一个共享的RNN编码器来建模意图和槽之间的关联，以此来实现槽与意图之间的联合，这可以被视为一个隐式的联合模型。Qin等人(2019)提出了一个堆叠传播模型 (Stack-Propagation)，以更好地利用意图语义信息来指导槽填充。上述模型可以被视为单一信息流向的联合模型。E等人(2019)提出了一个新颖的SF-ID网络，同时考虑到槽到意图和意图到槽的双向影响，为槽填充和意图识别提供了双向影响的机制。最近，多意图SLU逐渐受到关注，因为它在我们日常生活中具有重要的应用价值 (Gangadharaiah et al., 2019; Qin et al., 2000)，它可以处理包含多个意图的语句。Gangadharaiah 等人(2019)首次尝试提出了一个多任务框架，以联合建模多重意图识别和槽填充。Qin等人(2020)进一步提出了自适应交互框架 (AGIF)，以实现更细粒度的多意图信息集成。基于图注意网络和非自回归方式的思想，Qin等人(2021)提出了全局局部图交互网络 (GL-GIN)，该网络可以建模多个意图和话语中所有槽之间的槽依赖和交互，获得了先进的性能和效率。虽然该模型取得了较好的效果，然而，它在模型浅层交互中只是将多意图和槽填充信息简单的拼接，不能充分挖掘意图中对槽的填充具有重要的指导意义的信息，从而让训练效率和模型性能大打折扣，我们的模型主要基于GL-GIN上改进。此外，最新的研究还有Xing等人(2022)通过构建Co-guiding模型种实现两个任务之间的相互指导的新型模型，通过异构语义标签图实现多个意图识别和槽填充之间的相互指导。

为了进一步增强多意图与槽填充的关联性以及协同优化的效果，我们提出了一个多意图融合框架，其中核心结构是多意图融合层，该层包含了三种信息的融合过程，可以在意图识别和槽填充任务之间建立更稳定有效的连接。考虑到意图识别任务在联合效应中起着重要作用，我们还应用了意图强度层，以获得更丰富的意图在话语中的表示，从而提高整体框架的鲁棒性。我们在MixATIS(Hemphill et al., 1990)和MixSNIPS(Coucke et al., 2018)两个公共数据集上进行了实验，结果表明，我们的框架获得了最先进的性能。此外，由于我们充分利用了多意图信息，因此我们的模型更容易协调意图识别和槽填充任务，从而大大提高了训练效率。

3 问题定义

多意图识别. 给定输入序列 $x = (x_1, \dots, x_n)$ ，多意图识别可以被定义为一个多标签分类任务，其输出一个序列意图标签 $o_I = (o_1^I, \dots, o_m^I)$ ，其中 m 是给定话语中意图的数量， n 是话语的长度。

槽填充. 槽填充可以被视为一个序列标注任务，将输入话语 x 映射到一个槽输出序列 $o_S = (o_1^S, \dots, o_n^S)$ 。

4 模型

在本节中，我们将详细介绍我们的MIFFF模型。模型的体系结构如Figure 2所示。MIFFF模型由一个共享编码器、两个解码器和两次融合过程组成。

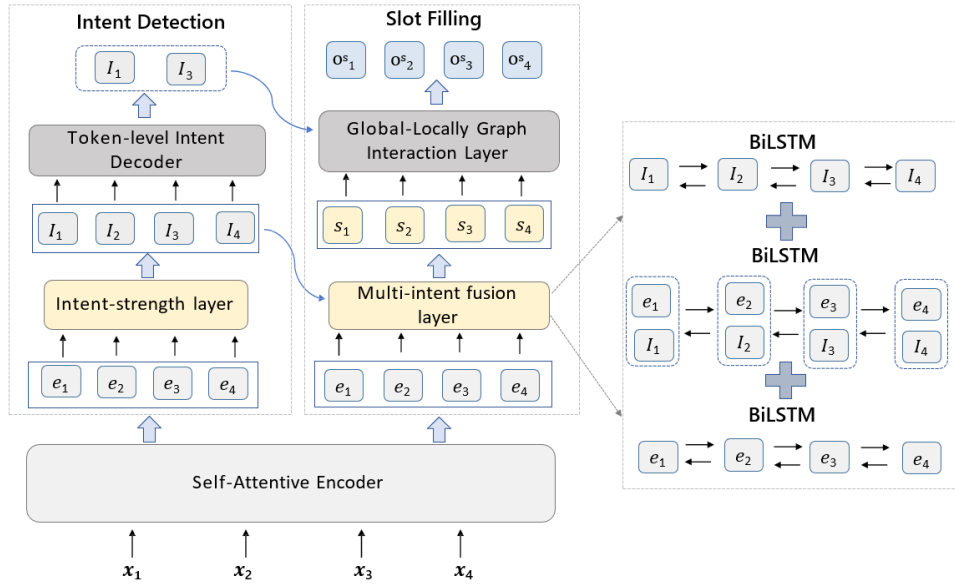


Figure 2: 模型架构和多意图融合过程

4.1 共享编码器

给定一个带有词元序列的句子 $\{t_1, t_2, \dots, t_n\}$, 输入嵌入层 φ^{emb} 将标记序列映射为嵌入序列 $X = x_1, x_2, \dots, x_n \in \mathbb{R}^{n \times d}$ (d 表示嵌入维度)。随着Qin 等人 (2019) 提出的方案, 具有双向LSTM (BiLSTM) 的自注意力编码器被用来捕获在词元顺序和上下文信息中的特征。BiLSTM (Chen et al., 2017) 通过使用 $h_i = \text{BiLSTM}(x_i)$ 生成上下文敏感的隐藏状态 $H = h_1, h_2, \dots, h_n$ 。受Vaswani 等人 (Shaw et al., 2018) 的启发, 词元矩阵表示上使用了自注意力机制。 $A = \text{Self-Attention}(H)$ 。将 H 和 A 并联成一个矩阵, 获取更丰富的语句编码信息:

$$E = H \parallel A. \quad (1)$$

4.2 意图强化层

我们构建了一个由带有残差连接 (He et al., 2016)的BiLSTM组成的意图强化层, 利用更丰富的意图表示, 最大限度地减少信息遗忘。我们将BiLSTM的输出和编码表示融合为意图表示:

$$\hat{h} = \text{BiLSTM}(E), \quad (2)$$

$$h = \text{BiLSTM}(\alpha \hat{h} + \beta E), \quad (3)$$

$$I = \sigma(W_I(\text{LeakyReLU}(W_h h + b_h)) + b_I), \quad (4)$$

其中 \hat{h} 和 h 用于增强特定任务的表征; $I = \{I_1, \dots, I_n\}$ 代表强化后的多意图隐藏层表示; α 和 β 是控制上下文意图信息强化超参数; σ 表示sigmoid激活函数; W_h 和 W_I 是可训练矩阵参数。

4.3 Token级别的意图识别解码器

我们使用token级别的多标签多意图识别 (Qin et al., 2021)来进一步提取有效的意图信息, 其中通过对所有预测的tokens投票来获得句子结果。句子的意图结果 o_k^I 可由下列公式获得:

$$o_k^I = \{o_k^I \mid \sum_{i=1}^n \mathbb{1}[I_{(i,k)} > 0.5] > n/2\}, \quad (5)$$

其中 $I_{(i,k)}$ 表示 i 对 o_k^I 的分类结果。当标签在所有 n 标记中获得一半以上的正向预测时, 我们就将其预测为语料意图, 这时我们可以捕捉到更细化的意图表示。

4.4 层次递进的多意图融合框架

4.4.1 多意图融合层

我们提出的多意图融合层主要由三个部分组成：首先，应用一个意图感知的BiLSTM来建模更详细的意图连接来丰富句子中的意图表示，同时我们将上下文编码E提供给一个槽感知的BiLSTM来增强其对于特定任务的表示。此外，我们利用一个BiLSTM来产生槽-意图隐藏表示，来加强多意图与槽信息的交互。最后，我们将来自三个不同维度的信息通过按元素相加，以获得更丰富的意图-槽信息表示，给槽填充提供更有效的指导，提高联合效果。融合过程可以定义为：

$$S'_1 = BiLSTM(E), \quad (6)$$

$$S'_2 = BiLSTM(I), \quad (7)$$

$$S'_3 = BiLSTM(I \parallel E), \quad (8)$$

$$S = S'_1 + S'_2 + S'_3, \quad (9)$$

其中 \parallel 代表了一个并联操作； $S = \{s_1, \dots, s_n\}$ 代表了多意图融合层之后的最终槽意识隐藏表示。

4.4.2 全局-局部交互层

我们将全局-局部图交互层应用到我们的框架中，以提取更加细颗粒的意图-槽交互信息 (Qin et al., 2021)。局部槽位感知图交互层是为了建立槽位间的依赖关系，缓解槽位不协调的问题。全局槽位-意图图交互层是为了实现句子级的意图-槽位交互，所有预测的多个意图和序列槽位都被连接，实现平行输出槽位序列的目的。他们的工作基于图注意网络(GAT)，将初始节点特征 $\tilde{H} = \{\tilde{h}_1, \dots, \tilde{h}_N\}$ ，旨在产生更抽象的表示 $\tilde{H}' = \{\tilde{h}'_1, \dots, \tilde{h}'_N\}$ 作为其输出。一个典型的GAT的注意机制可以概括如下：

$$\tilde{h}'_i = \parallel_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W_h^k \tilde{h}_j \right) \quad (10)$$

$$\alpha_{ij} = \frac{\exp \left(\text{LeakyReLU} \left(\mathbf{a}^\top [W_h \tilde{h}_i \parallel W_h \tilde{h}_j] \right) \right)}{\sum_{j' \in \mathcal{N}_i} \exp \left(\text{LeakyReLU} \left(\mathbf{a}^\top [W_h \tilde{h}_i \parallel W_h \tilde{h}_{j'}] \right) \right)} \quad (11)$$

其中， $W_h \in \mathbb{R}^{F' \times F}$ 和 $\mathbf{a} \in \mathbb{R}^{2F'}$ 是可训练的权重矩阵； \mathcal{N}_i 表示节点 i （包括 i ）的邻居； α_{ij} 是规范化的注意力系数， σ 代表非线性激活函数； K 是多头注意力数。

在通过线性层对每个图的邻居进行信息聚合后，我们能够获得更细颗粒的意图-槽交互信息表示。GAT中第 l 层的信息聚合过程可以表示为：

$$s_i^{l+1} = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W_l s_j^l \right), \quad (12)$$

其中， \mathcal{N}_i 是一组顶点，表示连接的槽-槽或意图-槽。在堆叠了 L 层之后，我们得到了上下文意图-槽感知的隐藏特征 $S^{L+1} = s_1^{L+1}, \dots, s_n^{L+1}$ 。

4.5 槽位预测

经过 L 层的传播，我们得到最终的槽位表示 s_i^{l+1} ，用于槽位预测，它可以被表述为为：

$$y_t^S = \text{softmax}(W_s s_i^{l+1}), \quad (13)$$

$$o_t^S = \text{argmax}(y_t^S), \quad (14)$$

其中 W_s 是一个可训练的参数， o_t^S 是预测的槽位， t 表示一个预测句子中的第 t 个词元， s_i^{l+1} 是经过 L 层的传播的聚合槽位信息。

4.6 协同训练

考虑到两个子任务之间的相关性，我们训练我们的模型，并联合更新参数，意图识别的目标可以表述为：

$$\mathcal{L}_1 = - \sum_{i=1}^n \sum_{j=1}^{n_I} \hat{y}_i^{(j,I)} \log \left(y_i^{(j,I)} \right), \quad (15)$$

其中 n_I 是意图的编号， $\hat{y}_i^{(j,I)}$ 是黄金意图标签。

同样地，槽填充的任务目标被表述为：

$$\mathcal{L}_2 = - \sum_{i=1}^n \sum_{j=1}^{n_S} \hat{y}_i^{(j,S)} \log \left(y_i^{(j,S)} \right), \quad (16)$$

其中 n_S 是槽的编号， $\hat{y}_i^{(j,S)}$ 是黄金槽标签。最终的联合目标是：

$$\mathcal{L} = \gamma \mathcal{L}_1 + (1 - \gamma) \mathcal{L}_2, \quad (17)$$

其中 γ 是一个用于平衡意图识别和槽填充任务的超参数。

5 实验

5.1 数据集

我们在两个公开的多意图SLU数据集MixATIS和MixSNIPS上进行了实验。MixATIS数据集是一个多意图数据集，由单意图数据集ATIS构建而成，用于评估自然语言理解模型的性能，它来自航空公司的查询，包括13,162个用于训练的语料，756个用于验证的语料和828个用于测试的语料。MixSNIPS数据集包含来自餐厅、酒店、电影等领域的查询，是由单意图数据集SNIPS构建而成的多意图数据集，它包括39,776、2,198和2,199个用于训练、验证和测试的语料。

5.2 实验设置

我们设定自我注意编码器的隐藏单元为256，丢弃率为0.4，LSTM隐藏单元的维度为256，意图嵌入层的维度为128，批量大小为16，多头注意力数为6，图注意力网络的层数为2， γ 为0.75， α 为0.8， β 为0.2，槽位预测解码器的维度为128。我们的模型以及复现的研究进展模型AGIF，GLGIN和Co-guiding都是使用在验证集上表现最好的模型，并在测试集上评估它的表现。我们用F1分数来评估槽填充的性能，用准确率来评估意图识别的性能，用总体准确率来评估语义解析的性能，它代表意图和槽位在语篇中全部被正确预测。我们所有的实验都是在RTX3090Ti上完成。

5.3 对比基线

我们将我们的模型与8个先进的基线模型进行比较，包括单意图SLU (Liu et al., 2016; Wang et al., 2018; E et al., 2019; Qin et al., 2019)和多意图SLU (Gangadharaiyah et al., 2019; Qin et al., 2020; Qin et al., 2021; Xing et al., 2022)。

- Attention BiRNN (Liu et al., 2016): 提出了一种基于注意力的神经网络模型，用于联合意图识别和槽填充，对于许多语音理解和对话系统影响深远。
- Bi-Model (Wang et al., 2018): 一种用于语音理解的模型，它考虑了意图和槽填充之间的交叉影响。
- SF-ID (E et al., 2019): 一种用于语音理解的模型，它是一种新颖的双向关联模型，用于联合意图识别和槽填充。
- Stack-Propagation (Qin et al., 2019): 一种基于堆栈传播的框架，用于意图识别和槽填充任务。

Model	MixATIS		
	Slot(F1)	Intent(Acc)	Overall(Acc)
Attention BiRNN (Liu et al., 2016)	86.4	74.6	39.1
Bi-Model (Wang et al., 2018)	83.9	70.3	34.4
SF-ID (E et al., 2019)	87.4	66.2	34.9
Stack-Propagation[sig] (Qin et al., 2019)	87.4	71.9	41.0
Joint Multiple ID-SF (Gangadharaiah et al., 2019)	84.6	73.4	36.1
AGIF (Qin et al., 2020)	86.9	72.2	39.2
GL-GIN (Qin et al., 2021)	87.2	76.0	42.5
Co-guiding (Xing et al., 2022)	86.53	74.03	43.35
MIFF	87.7*	77.2*	45.0*

Table 1: 在MixATIS的实验结果。带*的数字表示我们的模型对所有基线的改进在t检验下具有统计学意义, $p < 0.05$ 。

Model	MixSNIPS		
	Slot(F1)	Intent(Acc)	Overall(Acc)
Attention BiRNN (Liu et al., 2016)	89.4	95.4	59.5
Bi-Model (Wang et al., 2018)	90.7	95.6	63.4
SF-ID (E et al., 2019)	90.6	95.0	59.9
Stack-Propagation[sig] (Qin et al., 2019)	93.2	94.6	71.9
Joint Multiple ID-SF (Gangadharaiah et al., 2019)	90.6	95.1	62.9
AGIF (Qin et al., 2020)	93.8	95.1	72.7
GL-GIN (Qin et al., 2021)	93.9	95.5	72.5
Co-guiding (Xing et al., 2022)	93.8	95.1	72.7
MIFF	94.2*	95.8*	74.3*

Table 2: 在MixSNIPS中的实验结果。带*的数字表示我们的模型对所有基线的改进在t检验下具有统计学意义, $p < 0.05$ 。

- Joint Multiple ID-SF (Gangadharaiah et al., 2019): 一种多任务框架, 可以联合学习槽填充 (SF) 和多个意图识别 (ID)。
- AGIF (Qin et al., 2020): 一种自适应图交互框架, 用于联合多个意图识别和槽填充。
- GL-GIN (Qin et al., 2021): 一种快速准确的非自回归模型, 用于联合多个意图识别和槽填充。
- Co-guiding (Xing et al., 2022): 一种实现两个任务之间的相互指导的新型模型, 通过异构语义标签图实现多个意图识别和槽填充之间的相互指导。

5.4 主要结果

Table 1和2展示了我们的模型在MixATIS和MixSNIPS数据集上的实验结果, 我们有以下观察和分析:

(1) 在槽位填充任务上, 我们的框架在两个数据集上的F1得分超过了所有的强基线, 这表明我们的框架通过层次性多意图融合, 有效利用丰富的意图信息来指导槽位填充。

(2) 与GL-GIN相比, 我们的框架在MixATIS和MixSNIPS的总体精度上分别实现了2.5%和1.8%的提高。我们认为, 多意图融合框架能有效捕捉粗细颗粒的意图-槽信息表示, 提高总体任务的联合性能。

(3) 最重要的是, 我们的框架在所有评价指标上都达到了最先进水平, 展示了我们的框架在联合意图识别和槽填充任务的优越性, 为今后的工作留下了进一步探索融合机制的空间。

Model	MixATIS		
	Slot(F1)	Intent(Acc)	Overall(Acc)
MIFF	87.7	77.2	45.0
- w/o 意图强化层	88.0	76.4	44.0
- w/o 多意图融合层	87.7	76.0	42.6

Table 3: 在MixATIS数据集上的消融实验.

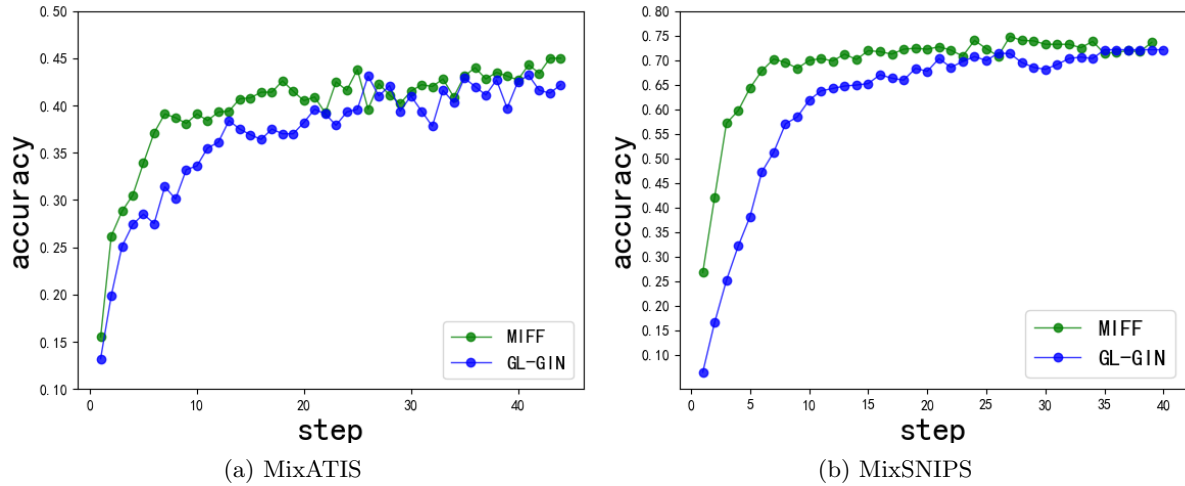


Figure 3: 训练效率的对比

5.5 进一步分析

在本节，我们主要在MixATIS数据集上研究分析了意图强化层和多意图融合层在我们整体模型框架的作用，结果如Table 3所示。

5.5.1 消去意图-强度层

为了检验意图强度层的有效性，我们将其替换为单一的BiLSTM层进行对照实验。实验结果显示，其在整体语义解析的准确性上降低了1.0%，在意图识别准确率上降低了0.8%。这表明我们提出的意图强化层对于丰富了多意图信息和缓解信息遗忘问题有重要作用，它可以通过集连多个维度的信息，获得更全面的意图表示信息，从而提高了多意图识别效能，同时提升整体语义解析的性能。

5.5.2 消去多意图融合层

为了进一步检验多意图融合层的有效性，我们将三种信息融合通道(意图-意图，槽-槽，槽-意图)替换为单一的槽-槽信息通道。结果显示，它的语义解析精度降低了2.4%。这意味着多意图融合对于联合意图识别和槽位填充非常重要。它可以为槽位任务提供更多有用的意图信息，提高联合意图识别和槽位填充之间的性能。

5.5.3 对槽填充任务的影响

从Table 3可以看出，在意图强化层和多意图融合层联合作用下，槽填充F1性能没有得到显著地提高，然而总体的语义解析性能得到改善，我们认为，我们提出的多意图融合框架更适用于提升总体的联合多意图识别和槽填充性能。

5.6 效率评估

我们在16个批次下评估了我们的框架和目前效率最高的非自回归框架(GL-GIN)之间的训练效率，如图3所示，在这两个数据集上，我们的框架相比GL-GIN有更卓越的训练效率。我们认为，多意图融合层可以更加准确地提供细粒度的意图表示信息，以加快槽的填充任务。同时我

们发现，我们的模型框架在多领域数据集MixSNIPS上的表现更加突出。我们认为，多意图融合框架能有效泛化到多领域的语料任务，这为今后的工作留下了进一步探索融合机制的空间。

6 总结与未来工作

在本文中，我们提出了一个用于联合多意图识别和槽位填充的多意图融合框架，它可以提取更丰富的意图信息来指导槽位填充任务，在槽位填充和意图识别之间建立更强的相关性，同时提高训练效率。我们在MixATIS和MixSNIPS两个公共数据集上进行了实验，结果表明，我们的模型在性能和效率方面均超过了当前最先进的方法，同时能够有效从单领域的MixATIS数据集泛化到多领域的MixSNIPS数据集上。

致谢

本文受到广东省自然科学基金(2021A1515011864)、国家自然科学基金(71472068)、广州市智慧农业重点实验室(201902010081)、广东省普通高校特色创新项目(2020KTSCX016)、国家级大学生创新训练计划项目(202210564069)的资助。

参考文献

- Gokhan Tur and Renato De Mori, "Spoken language understanding: Systems for extracting semantic information from speech" *Wiley, New York, 2011*.
- Steve Young, Milica Gašić, Blaise Thomson, and Jason D Williams, "Pomdp-based statistical spoken dialog systems: A review" *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1160–1179, 2013.
- Rashmi Gangadharaiah and Balakrishnan Narayanaswamy, "Joint multiple intent detection and slot labeling for goal-oriented dialog," *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019*, pp. 564–569
- Puyang Xu and Ruhi Sarikaya, "Convolutional neural network based triangular CRF for joint intent detection and slot filling," *In 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olo-mouc, Czech Republic, December 8-12, 2013*, pp. 78–83
- Hyun Kim, Jong-Hyeok Lee, and Seung-Hoon Na, "Predictor-estimator using multilevel task learning with stack propagation for neural quality estimation" *In Proceedings of the Second Conference on Machine Translation, 2017*, pp. 562–568.
- Libo Qin, Xiao Xu, Wanxiang Che, and Ting Liu, "AGIF: An adaptive graph-interactive framework for joint multiple intent detection and slot filling" *In Findings of the Association for Computational Linguistics: EMNLP 2020, Online, Nov. 2020*, pp. 1807–1816
- Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Li'o, and Yoshua Bengio, "Graph attention networks" *CoRR*, vol. abs/1710.10903, 2017
- Libo Qin, Fuxuan Wei, Tianbao Xie, Xiao Xu, Wanxiang Che, and Ting Liu, "GL-GIN: fast and accurate non-autoregressive model for joint multiple intent detection and slot filling" *In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pp. 178–188
- Charles T Hemphill, John J Godfrey, and George R Doddington, "The atis spoken language systems pilot corpus" *In Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990, 1990*
- Alice Coucke, Alaa Saade, Adrien Ball, Th eodore Bluche, Alexandre Caulier, David Leroy, Clement Doumouro, Thibault Gisselbrecht, Francesco Calta-girone, Thibaut Lavril, Ma el Primet, and Joseph Dureau, "Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces" *CoRR*, vol. abs/1805.10190, 2018.
- Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang, "Improving sentiment analysis via sentence type classification using bilstm-crf and cnn" *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017

- Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani, "Self-attention with relative position representations" *CoRR*, vol. abs/1803.02155, 2018.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition" *In 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27- 30, 2016*, pp. 770–778.
- Bing Liu and Ian Lane, "Attention-based recurrent neural network models for joint intent detection and slot filling" *In Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016)*, San Francisco, CA, USA, September 8-12, 2016, pp. 685–689
- Yu Wang, Yilin Shen, and Hongxia Jin, "A bi-model based RNN semantic frame parsing model for intent detection and slot filling" *In Proceedings of the 2018 Conference of the 16th North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2018)*, New Orleans, Louisiana, USA, June 1-6, Volume 2 (Short Papers), 2018, pp. 309–314.
- Haihong E, Peiqing Niu, and Zhongfu Chen, "A novel bi-directional interrelated model for joint intent detection and slot filling" *In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL 2019)*, Florence, Italy, July 28- August 2, Volume 1: Long Papers, 2019, pp. 5467–5471.
- Libo Qin, Wanxiang Che, and Yangming Li, "A stack-propagation framework with token-level intent detection for spoken language understanding" *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019)*, Hong Kong, China, November 3-7, 2019, pp. 2078–2087.
- Bowen Xing and Ivor W. Tsang "Co-guiding Net: Achieving Mutual Guidances between Multiple Intent Detection and Slot Filling via Heterogeneous Semantics-Label Graphs" *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022*, Abu Dhabi, United Arab Emirates, December 7-11, 2022
- Christian Raymond and Giuseppe Riccardi, "Generative and discriminative algorithms for spoken language understanding" *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association*, Antwerp, Belgium, August 27-31, 2007, pages= 1605–1608
- Patrick Haffner, Gokhan Tur, and Jerry H, "Optimizing SVMs for complex call classification" *In proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, Hong Kong, April 6-10, 2003, pp. 632–635.
- Robert E. Schapire and Yoram Singer "Booster: A Boosting-based System for Text Categorization" *Machine Learning*, 2000, 39(2/3):135–168.
- Kaisheng Yao, Baolin Peng, and Geoffrey Zweig, et al. "Recurrent conditional random field for language understanding" *In Proceedings of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, Florence, Italy, May 4-9, 2014, pp. 4077–4081.
- Kaisheng Yao, Baolin Peng, Yu Zhang, Dong Yu, Geoffrey Zweig and Yangyang Shi, "Spoken language understanding using long short-term memory neural networks" *in Proceedings of the 2014 IEEE Workshop on Spoken Language Technology (SLT 2014)*, South Lake Tahoe, NV, USA, December 7-10, 2014, pp. 189–194.
- Rashmi Gangadharaiah and Balakrishnan Narayanaswamy, "Joint multiple intent detection and slot labeling for goal-oriented dialog" *in Proceedings of the 17th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, Minneapolis, MN, USA, June 2-7, Volume 1 (Long and Short Papers), 2019, pp. 564–569.
- Libo Qin, Xiao Xu, Wanxiang Che, et al. "Towards fine-grained transfer: An adaptive graph-interactive framework for joint multiple intent detection and slot filling" *in Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November, 2020*, vol. EMNLP 2020 of Findings of ACL, pp. 1807– 1816.