主题感知的多意图识别与槽位填充联合建模方法

罗晶[1,2,3]，王炜华[1,2,3,*]，曹越[1,2,3]，飞龙[1,2,3]，高光来[1,2,3]
1.内蒙古大学计算机学院，内蒙古，呼和浩特，010021
2.蒙古文智能信息处理技术国家地方联合工程研究中心，内蒙古，呼和浩特，010021
3.内蒙古自治区蒙古文信息处理技术重点实验室，内蒙古，呼和浩特，010021
297317497@qq.com，wangwh@imu.edu.cn,caoyueimu@163.com
csfeilong@imu.edu.cn，csggl@imu.edu.cn

摘要

意图识别与槽位填充是口语理解中的两个子任务，联合建模这两项任务能够利用共享特征提升任务间的协同建模效果。然而，现有方法普遍缺乏对句子主题语义的显式建模，难以捕捉更充分的全局语义信息，尤其在多意图场景下系统建模性能下降严重。为缓解上述问题，本文提出了一种主题感知的意图识别与槽位填充联合建模方法，该方法构造了主题提取模块以学习句子主题分布表示，结合主题引导的意图和槽位表示增强网络插入主题信息，使得模型在识别句子意图和填充槽位过程中能够显式建模主题信息。实验结果表明，本文所提出方法在多意图公开数据集MixATIS和MixSNIPS上分别获得了50.9%和84.8%的整体准确率，相较多个基线模型取得了更优的性能表现。

关键词： 意图识别；槽位填充；自然语言理解；对话系统

A Topic-Aware Joint Modeling Method for Intent Detection and Slot Filling

Jing Luo[1,2,3], Weihua Wang[1,2,3,*], Yue Cao[1,2,3], Feilong Bao[1,2,3], Guanglai Gao[1,2,3]
1.College of Computer Science, Inner Mongolia University, Hohhot 010021, China
2.National and Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian, Hohhot 010021, China
3.Inner Mongolia Key Laboratory of Mongolian Information Processing Technology, Hohhot , China
297317497@qq.com，wangwh@imu.edu.cn,caoyueimu@163.com
csfeilong@imu.edu.cn，csggl@imu.edu.cn

Abstract

Intent recognition and slot filling are two sub-tasks of spoken language understanding. Jointly modeling of these tasks can enhance their collaborative performance by sharing features. However, existing methods generally lack explicit modeling of sentence-level topic semantics, making it difficult to capture more comprehensive global semantic information, particularly in multi-intent scenarios, where they exhibit certain limitations. To address these issues, this paper proposes a topic-aware joint modeling approach for intent recognition and slot filling (TA-IDSF). The method constructs a topic extraction module to obtain sentence-level topic distribution representations, and introduces topic information into both tasks using topic-guided intent and slot representation enhancement networks, applying cross-attention and concatenation techniques, respectively. This enables the model to leverage topic information to provide more effective guidance for intent recognition and slot filling tasks. Experimental results on the multi-intent public datasets MixATIS and MixSNIPS show that the proposed method

©2025 中国计算语言学大会
根据《Creative Commons Attribution 4.0 International License》许可出版

第二十四届中国计算语言学大会论文集，第525页-第536页，济南，中国，2025年8月11日至14日。

achieves overall accuracy rates of 50.9% and 84.8%, respectively, outperforming several baseline models.

Keywords: Intent Detection , Slot Filling , Natural Language Understanding , Dialogue Systems

1 引言

口语理解（Spoken Language Understanding, SLU）(Tur and De Mori, 2011; Young et al., 2013)是任务型对话系统的核心模块，主要包含意图识别和槽位填充两个子任务。意图识别是判断用户所表达的意图类别，该类别通常来自一个预定义的候选集合，它是一个文本分类 (Kowsari et al., 2019)问题。槽位填充本质是把用户输入语句中的时间、地点等重要信息标注上定义好的槽位名称，它是一种序列标注任务 (Nguyen and Guo, 2007)。这两个任务相互关联，意图识别为槽位填充提供语义方向，槽位填充则为意图识别提供细节语义支撑，两者协同完成对用户语句的完整理解。例如，句子“what is the forecast now in kansas and also give two out of 6 to current book”，包含两个意图，分别为：“GetWeather”（意图为：获取天气）和“RateBook”（意图为：对图书评分），槽位填充的标注结果为：“`O O O O B-timeRange O B-state O O O B-rating_value O O B-best_rating O B-object_select B-object_type` ”。由于意图识别任务与槽位填充任务紧密相关，SLU模型通常采用联合建模的方式来学习它们之间的关系。然而，传统研究多聚焦于单意图场景，难以应对复杂多意图表达的现实需求。随着多意图SLU在实际应用中的重要性日益凸显，相关研究逐渐兴起。Xu (2013)等人和Kim (2017)等人尝试了多意图SLU的建模，他们的工作主要关注意图识别，忽略了槽位填充任务的协同优化。Gangadharaiah和Narayanaswamy (2019)提出了一种多任务框架，尝试同时处理多意图识别与槽位填充。Qin (2020)等人则进一步设计了一种自适应交互框架（AGIF），通过细粒度的方式整合多意图信息。此外，基于图注意力机制和非自回归解码思想，Qin (2021)等人提出了全局-局部图交互网络（GL-GIN），该模型通过捕捉多意图及语句中槽位间的依赖关系，显著提升了性能。Cai (2022)等人引入BERT并结合显式的槽位-意图映射机制，通过强化意图与槽位的双向交互，显著提升了多意图场景下的语义理解能力。然而，现有方法缺乏对全局语义结构的显式建模，在复杂多意图场景仍有一定局限性。

为了在多意图场景下更全面地提升模型的全局语义建模能力，我们提出一种主题感知的意图识别与槽位填充联合建模方法，构建了主题提取模块，利用神经主题提取方法获取的主题分布信息作为语义引导，具体而言，本文首先通过主题提取模块从输入文本中获取句子级主题分布表示，并将其定义为全局语义信息约束。随后，围绕意图识别与槽位填充两个子任务设计差异化融合策略：在意图识别中引入基于交叉注意力的融合方式，以提升模型对整句主题的理解能力；在槽位填充中采用拼接的融合方式，在引入全局语义信息的同时保留局部边界特征的表达优势。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，我们的模型获得了最先进的性能。

本文贡献如下：（1）构建了一个主题提取模块，其引入的全局语义信息可以有效提升意图识别与槽位填充两个任务的语义建模能力。（2）为了更好地利用主题信息，我们针对意图识别任务和槽位填充任务设计了差异化的主题引导机制。（3）本文提出的主题感知的意图识别与槽位填充联合建模方法在MixATIS与MixSNIPS两个公开多意图数据集上均取得了有竞争力的实验结果。

2 相关工作

2.1 意图识别任务

意图识别任务的核心目标在于将用户输入正确分类至预定义的意图类别，其发展大致可分为两个阶段。2015年以前，研究者主要采用传统方法，包括基于规则模板的方法 (Marslen-Wilson and Tyler, 1980)和基于统计学习的方法 (Genkin et al., 2007)。随着数据规模的持续增长与任务复杂度的提升，传统方法的局限性逐渐显现。自2015年起，随着深度学习技术的迅速发展，研究者逐步转向基于神经网络的建模方法，以替代传统的统计学习手段。这类方法能够

(c) 2025 中国中文信息学会计算语言学专业委员会

自动学习文本的语义表示，在意图识别任务中显著提升了性能。在早期研究中，Chen (2015)等人提出了一种基于卷积神经网络的模型，用于处理多标签分类问题，为意图识别提供了新的思路。随后，Lai (2015)等人提出一种无需人工设计特征的循环卷积神经网络(Recurrent Convolutional Neural Network, RCNN)，不仅减少了对人工特征工程的依赖，还能够处理更复杂的语言现象。此外，BERT (Khan and Meenai, 2021)通过预训练与微调机制在意图识别任务中展现出卓越性能，并为后续意图识别与槽位填充的联合建模提供了有力支撑与启发。

2.2 槽位填充任务

槽位填充任务的核心目标在于识别句子中的时间、地点等重要语义成分，并为其分配对应的槽位标签。该任务通常被建模为一个序列标注问题。传统的槽位填充方法主要包括基于规则的方法与基于统计机器学习的方法。近年来，深度学习技术显著提升了槽位填充的效果。研究者们尝试将循环神经网络(Recurrent Neural Network, RNN) (Zaremba et al., 2014)及其变体应用于槽位填充任务。例如，Mesnil (2013)等人将RNN引入槽位填充，通过建模时间依赖关系，取得了优于传统CRF模型的效果。Vu (2016)等人采用双向循环神经网络(Bi-directional Recurrent Neural Network, BIRNN)并结合排序损失函数，进一步提升了模型性能。Kurata (2016)等人则利用长短期记忆网络(Long Short-Term Memory, LSTM)来捕捉标签依赖性和上下文信息，显著提高了槽位填充的准确性。此外，徐梓翔 (2017)等人提出的BiLSTM-CRF模型，结合了双向LSTM和CRF的优势，在公开数据集上取得了显著提升。

2.3 意图识别与槽位填充联合建模

早期的自然语言理解模型通常采用管道式建模方法，将意图识别与槽位填充视为独立任务分别训练与推理。然而，这种方法易导致误差传播。为此，联合建模方法逐渐成为研究热点。Qin (2019)等人提出了一种基于多阶段预测的框架，利用堆栈结构逐步融合意图信息指导槽位填充。Gangadharaiah (2019)等人通过将多意图识别建模为多标签分类任务，实现意图识别与槽位填充的信息共享，增强联合建模能力。Qin (2020)等人提出了一种基于图注意力网络的动态交互模块，利用细粒度多意图标签构建动态图，增强意图与槽位间的语义关联，引导槽位填充任务。Qin (2021)等人进一步设计了一种结合局部与全局图交互的口语理解模型，通过建模槽位间依赖与意图-槽位交互，并引入非自回归机制提升推理效率与信息传递效果。Chen (2022)等人引入顺序连接解码器与自蒸馏机制，通过多实例学习将多意图识别建模为弱监督任务，利用词级意图信息协同指导槽位填充，表现出较强的竞争力。Cai (2022)等人则结合BERT与显式槽位-意图映射机制，强化意图与槽位的双向交互，显著提升多意图场景下的语义理解效果。Nguyen 等人 (2023)提出了一种双向联合模型，该模型引入了一种新颖的训练方法，通过监督对比学习与自蒸馏机制来优化所提出的联合模型。Zhu等人 (2024)提出了一种场景感知的标签图交互框架，通过对用户语句场景的自动分类并维护场景特定的标签共现矩阵，显式建模标签间的动态共现关系。

现有研究多从结构设计与任务协同两方面提升意图与槽位之间的语义交互能力，虽取得显著进展，但在全局语义显式建模方面仍存在不足。本文引入主题信息以提供全局语义辅助联合建模。

3 问题定义

假设用户当前输入语句为$U = \{w_1, w_2, ..., w_n\}$，其中$n$ 表示该语句中包含的词数量。

意图识别任务： 在预设的意图标签集合$\mathrm{I_{all}} = \{i_1, i_2, ..., i_m\}$ 中，识别出用户输入语句的意图标签$I \in \mathcal{I}_{\mathrm{all}}$，其中$m$ 为意图标签总数量。

槽位填充任务： 从用户输入的字符序列U中识别出每个字符对应的槽位标签序列$\mathcal{S} = \{\mathrm{s_1}, \mathrm{s_2}, ..., \mathrm{s_n}\}$。

因此，给定用户输入语句U，意图识别与槽位填充联合建模的目标可定义为训练一个模型P，该模型能够同时识别用户输入文本的意图类别I以及槽位标签序列S。如公式1 所示:

$$I, S = P(U|\theta) \tag{1}$$

其中，θ 表示模型P的可训练参数。

(c) 2025 中国中文信息学会计算语言学专业委员会

4 主题感知的意图识别与槽位填充联合模型方法

本节模型基于多任务学习框架构造，整体架构如图 1所示。模型主要由以下五个模块组成，分别是：文本编码模块、主题提取模块、基于主题引导的意图表示增强网络、基于主题引导的槽位表示增强网络以及解码器模块。

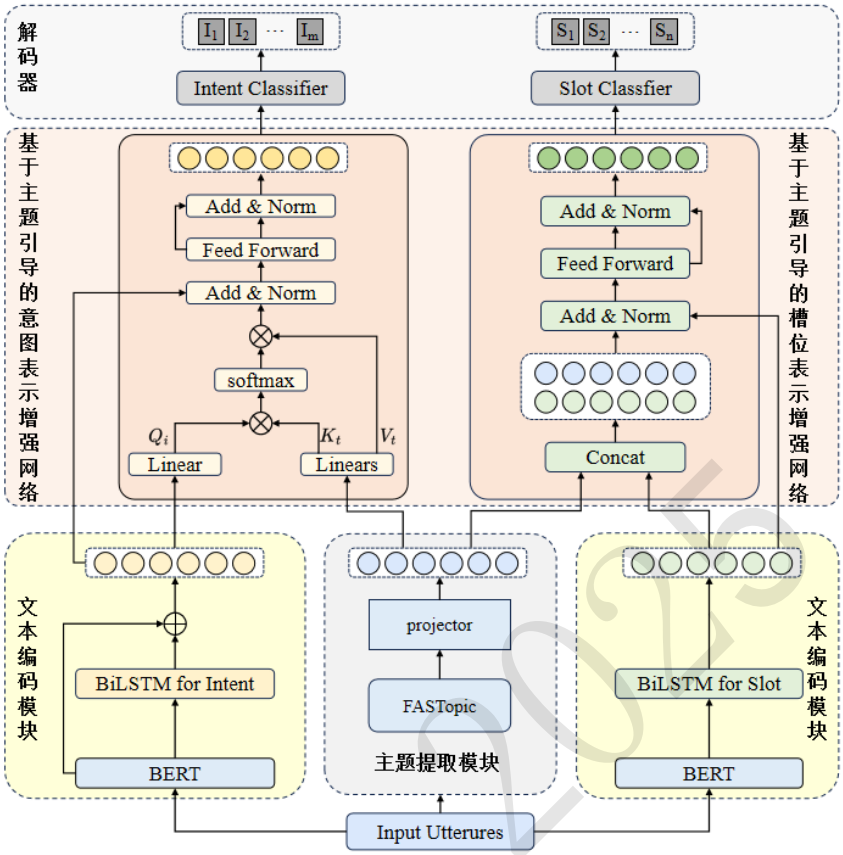


Figure 1: 主题感知的意图识别与槽位填充联合建模方法架构图

4.1 文本编码模块

给定一个输入序列$U = \{w_1, w_2, ..., w_n\}$，其中$n$ 是句子的长度，通过BERT得到了整个句子的向量表示$H_{[CLS]}$ 以及对应的上下文表示H，如公式2 所示。

$$H = \{h_1, h_2, ..., h_n\} \tag{2}$$

其中，h_i 表示序列中第i 个词汇的上下文嵌入表示。

本文将BERT输出的上下文H表示传入一个意图感知的BiLSTM网络得到H_{intent}，进一步优化了序列表征能力，考虑到深层网络在语义传播过程中可能出现的信息衰减问题，本文在BiLSTM输出后加入残差连接，将BERT原始的上下文表示H 与BiLSTM的输出特征H_{intent}进行融合，以增强模型对整句语义的保留能力。意图识别任务的编码特征H_{intent}的计算如公式3和4 所示：

$$H_{\text{intent}} = \text{BiLSTM}_{\text{intent}}(H) \tag{3}$$

$$H_{\text{intent}} = H + H_{\text{intent}} \tag{4}$$

同样地，将BERT输出的语义表示H输入至一个槽位感知的BiLSTM网络，以补充其局部依赖建模能力。槽位填充任务的编码H_{slot}表示如公式5 所示：

$$H_{\text{slot}} = \text{BiLSTM}_{\text{slot}}(H) \tag{5}$$

第二十四届中国计算语言学大会论文集，第525页-第536页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

4.2 主题提取模块

该模块引入了神经主题建模方法FASTopic (Wu et al., 2024)，该方法有效缓解了传统主题建模方法中常见的关系偏差问题。FASTopic所学习到的主题概率向量以词汇表的概率分布形式表示，蕴含了来自语料整体的语义背景信息。借助该方法，模块能够从输入文本中生成文档-主题分布矩阵，该矩阵刻画了不同句子在各主题上的语义偏好，本文将由此获得的句子级主题分布表示定义为“全局语义信息”，与局部上下文建模相比，该信息提供了对整体语境的宏观语义感知。具体来说，对于输入序列$U = \{w_1, w_2, \ldots, w_n\}$（其中$n$ 是句子的长度），本文使用FASTopic主题建模方法得到每个主题对应的关键词集合以及每条输入句子的主题分布。本文将第i 条句子的主题分布表示为向量$t_i \in \mathbb{R}^k$，其中k 为设定的主题数，表示该句子在k 个主题上的归一化概率分布，满足$\sum_{j=1}^{k} t_{ij} = 1$。$t_{ij}$ 表示第i 个句子在第j 个主题上的概率得分。将批次中所有句子的主题分布向量按行堆叠，得到文档-主题分布矩阵$T \in \mathbb{R}^{b\times k}$，其中$b$ 表示批次大小，T 的第i 行即为向量t_i。该矩阵整体反映了当前批次中所有句子在各个主题上的分布情况。

为适配后续的特征融合操作，本文首先将文档-主题分布矩阵T 在第二维插入序列长度维度进行扩展，以匹配下游任务的上下文特征结构。随后通过线性变换将文档-主题分布矩阵T投影至隐藏空间，如公式6 所示：

$$T = \text{Projection Layer}(T) = TW_{\text{t}} + b_{\text{t}} \tag{6}$$

其中，$W_z \in \mathbb{R}^{K\times d_h}$、$b_z \in \mathbb{R}^{d_h}$ 为可训练参数。

4.3 基于主题引导的意图表示增强网络

为进一步提升模型对意图识别任务的建模能力，我们设计了基于主题引导的意图表示增强网络。该网络通过交叉注意力将全局语义信息引入意图识别任务中，以强化模型对话语整体语义的理解能力，并采用基于Transformer的多层编码器进一步编码。

具体而言，本文以意图特征H_{intent} 作为查询（Query），主题特征T 作为键（Key）和值（Value），计算交叉注意力分数，以实现主题特征对意图特征的融合。首先，对意图特征H_{intent}和主题特征T分别应用线性变换得到查询矩阵Q_{intent}、键矩阵K_{topic} 和值矩阵V_{topic}。具体计算方式如下：

$$Q_{\text{intent}} = H_{\text{intent}} W_Q^{\text{intent}} \tag{7}$$

$$K_{\text{topic}} = TW_K^{\text{topic}} \tag{8}$$

$$V_{\text{topic}} = TW_V^{\text{topic}} \tag{9}$$

其中，$W_Q^{\text{intent}}, W_K^{\text{topic}}, W_V^{\text{topic}}$ 为可训练的线性变换权重矩阵。

随后，使用查询矩阵Q_{intent}与键矩阵K_{topic}计算注意力权重矩阵A_{it}，并通过softmax对权重进行归一化，如公式10 所示：

$$A_{it} = \text{softmax}\left(\frac{Q_{\text{intent}}(K_{\text{topic}})^{\text{T}}}{\sqrt{d_k}}\right) \tag{10}$$

其中，T表示矩阵转置操作（此处为避免与主题增强特征T混淆），d_k 是一个标准的缩放因子。

接着，本文使用注意力权重矩阵A_{it}与值矩阵V_{topic}进行加权相乘，获得经过主题增强特征$H_{\text{intent}}^{\text{topic}}$强化后的意图增强特征，如公式11 所示：

$$H_{\text{intent}}^{\text{topic}} = A_{it} V_{\text{topic}} \tag{11}$$

为了保持原始意图特征的稳定性，本文引入残差连接与层归一化操作，将原始意图特征H_{intent}与意图增强特征$H_{\text{intent}}^{\text{topic}}$进行融合，得到初步融合后的特征表示，如公式12 所示：

$$H_{\text{intent}}^{\text{fusion}} = \text{LayerNorm}(H_{\text{intent}} + H_{\text{intent}}^{\text{topic}}) \tag{12}$$

(c) 2025 中国中文信息学会计算语言学专业委员会

最后，为进一步增强模型的非线性表达能力，本文使用前馈神经网络在其输出后进行残差连接与层归一化操作，最终获得意图增强特征$H_{\text{intent}}^{\text{fusion}}$，如公式13 和14 所示：

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{13}$$

$$H_{\text{intent}}^{\text{fusion}} = \text{LayerNorm}\left(H_{\text{intent}}^{\text{fusion}} + \text{FFN}\left(H_{\text{intent}}^{\text{fusion}}\right)\right) \tag{14}$$

其中，W_1, W_2 分别为前馈网络中的投影矩阵，b_1, b_2 为对应的偏置项。

通过上述融合过程，模型能够有效地引入主题信息，增强意图特征的表达能力，从而提升后续意图分类的准确性。

4.4 基于主题引导的槽位表示增强网络

为进一步提升模型对槽位填充任务的建模能力，本文设计了基于主题引导的槽位表示增强网络。不同于意图识别任务对句级语义的强依赖，槽位填充更侧重于局部上下文建模，对全局语义的依赖程度相对较低。因此，为避免对token表示产生重构性扰动，进而削弱模型对局部边界的建模能力，本文以拼接和线性映射的方式将主题信息引入到槽位特征中，并采用基于Transformer的多层编码器进行编码。该方法能够为每个词汇提供全局语义引导，在提升槽位预测准确性的同时，最大程度保持原始token表达的完整性。

具体而言，本文将槽位特征 H_{slot} 与主题特征 T 在隐藏维度上进行拼接，以融合槽位与主题信息，拼接后通过线性变换将特征映射回原始表示空间，如公式15和公式16 所示：

$$H_{\text{slot}}^{\text{topic}} = \text{Concat}(H_{\text{slot}}, T) \tag{15}$$

$$H_{\text{slot}}^{\text{topic}} = H_{\text{slot}}^{\text{topic}} W + b \tag{16}$$

为了避免引入主题信息后，重要局部语义被稀释，本文引入残差连接与层归一化操作，将原始槽位特征H_{slot}与槽位增强特征H_{slot}^{topic}进行融合，得到初步融合后的特征表示，如公式17 所示：

$$H_{\text{slot}}^{\text{fusion}} = \text{LayerNorm}\left(H_{\text{slot}} + H_{\text{slot}}^{\text{topic}}\right) \tag{17}$$

接着，为提升模型对融合信息的非线性建模能力，本文使用了前馈神经网络，在其输出后进行残差连接与层归一化操作，最终获得槽位增强特征，如公式18 所示：

$$H_{\text{slot}}^{\text{fusion}} = \text{LayerNorm}\left(H_{\text{slot}}^{\text{fusion}} + \text{FFN}\left(H_{\text{slot}}^{\text{fusion}}\right)\right) \tag{18}$$

通过上述过程，模型能够有效地融合主题信息，增强槽位特征的表达能力，从而提升后续槽位填充的性能。

4.5 解码器

在完成主题增强特征分别与意图特征和槽位特征的融合后，本文针对意图识别与槽位填充两项任务分别使用了不同类型的解码结构。

对于意图识别任务，本节对意图增强特征$H_{\text{intent}}^{\text{fusion}}$进行平均池化操作，得到句子级表示特征：

$$h_{\text{intent}} = \text{AvgPooling}(H_{\text{intent}}^{\text{fused}}) \tag{19}$$

随后，将该特征h_{intent}送入一个全连接分类器，结合softmax与argmax操作输出最终意图类别：

$$\hat{y}_{\text{intent}} = \text{argmax}\left(\text{softmax}(W_{\text{intent}} h_{\text{intent}} + b_{\text{intent}})\right) \tag{20}$$

其中，W_{intent} 为权重矩阵，b_{intent} 为偏置项，$\hat{y}_{\text{intent}}$ 为预测的意图标签。

为优化意图识别任务的分类性能，本文采用交叉熵损失函数（Cross Entropy Loss）作为目标函数，对预测标签$\hat{y}_{\text{intent}}$ 与真实标签y_{intent}之间的匹配程度进行度量。具体损失函数如公式21 所示：

$$\mathcal{L}_{\text{intent}} = -\sum_{c=1}^{C} y_{\text{intent}}^{(c)} \log\left(\hat{y}_{\text{intent}}^{(c)}\right) \tag{21}$$

第二十四届中国计算语言学大会论文集，第525页-第536页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

其中，$y_{\text{intent}}^{(c)}$ 为真实意图标签在第c 类上的one-hot表示，$\hat{y}_{\text{intent}}^{(c)}$ 为预测的意图类别在第c 类上的概率，C 为意图类别总数。

对于槽位填充任务，本文首先将融合后的槽位特征$H_{\text{slot}}^{\text{fusion}}$输入至线性映射层，得到每个token在槽位标签空间中的得分向量s_j:

$$s_j = W_{\text{slot}} h_{\text{slot},j}^{\text{fusion}} + b_{\text{slot}} \tag{22}$$

其中，W_{slot} 为映射权重矩阵，$h_{\text{slot},j}^{\text{fusion}}$ 表示第j 个token的融合表示，b_{slot} 为偏置项。

随后，本节将整个序列中所有位置的得分向量$\{s_1, s_2, ..., s_n\}$ 输入条件随机场（CRF）解码器中，建模标签之间的转移关系，实现对整个槽位标签序列的全局最优预测:

$$\hat{y}_{\text{slot}} = \text{CRF}(s_1, s_2, ..., s_n) \tag{23}$$

其中，W_{slot} 和b_{slot} 分别表示线性映射的权重矩阵和偏置项。

为优化槽位填充任务的性能，本文采用CRF的负对数似然损失函数，衡量预测标签序列$\hat{y}_{\text{slot}}$ 与真实标签序列y_{slot} 之间的匹配程度:

$$\mathcal{L}_{\text{slot}} = -\log p(y_{\text{slot}}|s_1, s_2, ..., s_n) \tag{24}$$

其中，y_{slot} 为真实槽位标签序列，$p(y_{\text{slot}}|\cdot)$ 为给定模型预测的得分向量后，真实槽位标签序列的条件概率，由CRF解码器计算得到。

最后，本文定义的意图识别与槽位填充联合训练目标函数，如公式25 所示:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{intent}} + \alpha \mathcal{L}_{\text{slot}} \tag{25}$$

其中，λ 是超参数，用来平衡意图损失和槽填充损失。

5 实验

5.1 数据集

我们在两个英文多意图公开数据集MixATIS和MixSNIPS (Qin et al., 2020)上开展实验。这两个数据集在ATIS与SNIPS数据集基础上构建，旨在模拟多意图场景。其中，ATIS数据集 (Hemphill et al., 1990)来源于旅客和航空预订系统之间的语音交互，涵盖21种意图类型和120类槽位标签。SNIPS数据集 (Coucke et al., 2018)来源于SNIPS个人语音助手，涵盖7种意图类型和72类槽位标签。模型的输入为文本模态的句子。为保证公平比较，我们采用了与各基线方法一致的数据划分方式。MixATIS数据集包含18000条训练集，1000条开发集和1000条测试集。MixSNIPS数据集包含45000条训练集，2500条开发集和2500条测试集。这两个数据集中，拥有1-3 个意图标签的样本比例为3:5:2。

5.2 实验设置

我们使用的超参数配置如下：BERT模型中的Transformer层数为12层，隐藏神经元个数为768。训练时采用的批处理大小为16，训练轮数设为20，学习率为0.0001，Dropout率为0.5，优化器选择AdamW。

5.3 基线方法

为了验证本文提出的TA-IDSF模型的有效性，本节将该模型与以下基线方法进行了比较。

- Slot-Gated (Goo et al., 2018): 该方法提出一种基于全注意力机制的联合建模方式，通过意图与槽位注意力计算捕获任务关联，利用槽门控结构以意图信息引导槽位填充任务，实现单向交互，提升预测性能。

- Stack-Propagation (Qin et al., 2019): 该方法是基于多阶段预测的口语理解框架，用堆栈融合意图信息指导槽位填充，分阶段执行词级意图预测与序列更新，降低错误累积，提升稳健性与泛化能力。

(c) 2025 中国中文信息学会计算语言学专业委员会

- JointMultiple ID-SF (Gangadharaiah, 2019): 该方法联合建模多意图识别与槽位填充，将多意图识别建模为多标签分类，实现信息共享，提升联合建模能力。
- AGIF (Qin et al., 2020): 该方法基于图注意力网络的动态交互模块，用多意图标签信息构建动态图，引导槽位填充，捕获意图与槽位关联，提升语义理解性能。
- GL-GIN (Qin et al., 2021): 该方法通过局部与全局图交互模块建模槽位依赖及意图-槽位交互，采用非自回归策略提升推理速度，实现高效信息传递。
- SDJN (Chen et al., 2022): 该方法引入顺序连接解码器与自蒸馏机制，增强意图-槽位交互，利用多实例学习建模多意图识别为弱监督任务，词级意图信息指导槽位填充，表现竞争力。
- SLIM (Cai et al., 2022): 该方法通过引入BERT结合显式槽位-意图映射机制，充分利用槽位信息指导意图识别，并在联合建模任务中增强意图与槽位的双向交互，显著提升了多意图场景下的语义理解效果。
- MIFF (Yin et al., 2023): 该方法通过提出了一个多意图融合框架，其核心是多意图融合层，包含了三种信息的融合过程，同时提出了意图强化层，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。

Model	MixATIS			MixSNIPS		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
Slot-Gated	88.1	65.7	38.9	87.8	96.0	56.5
Stack-Propagation	86.6	76.0	42.8	93.9	96.4	75.5
JointMultiple ID-SF	87.5	73.1	38.1	91.0	95.7	66.6
AGIF	88.1	75.8	44.5	94.5	96.5	76.4
GL-GIN	88.3	76.3	43.5	94.9	95.6	75.4
SDJN	88.2	77.1	44.6	94.4	96.5	75.7
SLIM	88.5	78.3	47.6	96.5	97.2	84.0
MIFF	87.7	77.2	45.0	94.2	95.8	74.3
TA-IDSF	**88.9**	**81.9**	**50.9**	**96.9**	**97.5**	**84.8**

Table 1: 本文模型（TA-IDSF）与现有模型在MixATIS 和MixSNIPS 数据集上的对比实验结果

5.4 主要实验结果

Table 1展示了我们的模型在MixATIS和MixSINPS数据集上的实验结果，TA-IDSF模型在MixATIS数据集和MixSNIPS数据集上的所有评估指标上均取得了最优性能，表现出明显的优势。从实验结果可以看出：

在意图识别方面，本文模型在两个数据集上的准确率分别达到81.9%和97.5%，均超越表现最好的基线SLIM (Cai et al., 2022)，提升幅度最高可达3.6个百分点；在槽位填充任务中，本文模型在两个数据集上分别取得了88.9%和96.9%的$F1$值，保持领先地位；在整体准确率方面，本文模型在MixATIS数据集和MixSNIPS数据集上分别达到50.9%和84.8%，高于基线SLIM（分别为47.6%和84.0%）。

上述性能优势的取得，归因于本文模型在结构设计与建模机制方面的多项改进。首先，本文构建了主题提取模块获取全局语义信息，并分别融入至意图识别和槽位填充两个子任务的表示空间，弥补了传统模型在面向多意图任务时全局语义建模不足的缺点。其次，模型通过差异化的引导机制分别建构意图-主题和槽位-主题之间的融合方式，使模型能够更精准地区分并建模多意图之间以及槽位之间的语义边界。最后，模型整体采用多任务联合优化目标函数，不仅提升了两个子任务的独立表现，也进一步促进了意图与槽位之间语义信息的共享与对齐，增强了模型的泛化能力和鲁棒性。

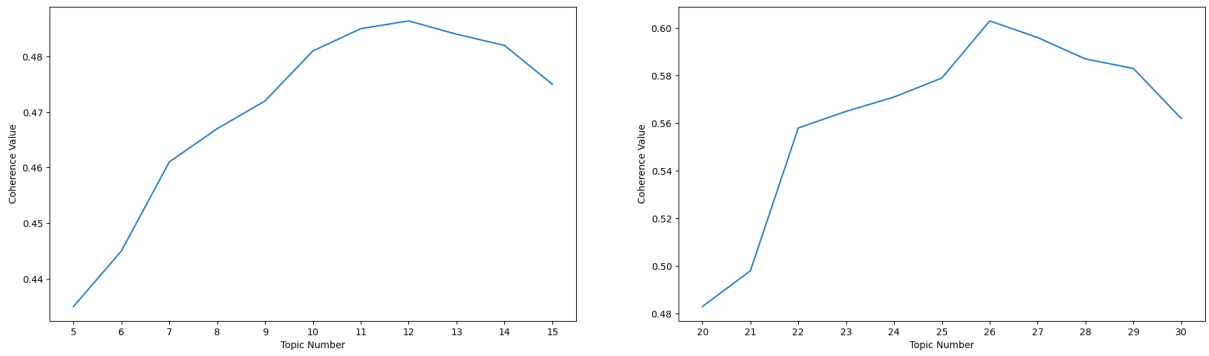
第二十四届中国计算语言学大会论文集，第525页–第536页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

5.5 进一步实验分析

5.5.1 主题数寻优实验

TA-IDSF的主题提取模块依赖于预设的主题数K来生成句子对应主题的概率分布，因此合理地设定主题数对于提升主题质量及意图识别与槽位填充任务的建模效果具有重要意义。我们分别在$[5, 6, \ldots, 15]$和$[20, 21, \ldots, 30]$的区间内搜索合适的主题数，并基于主题一致性指标，选择得分最高的主题数作为模型最终配置。实验结果如图 2 所示。



(a) MixSNIPS 数据集　　(b) MixATIS 数据集

Figure 2: 不同数据集在主题数变化下的模型表现

由图 2a和图 2b可知，当主题数分别为12（MixSNIPS）与26（MixATIS）时，主题一致性指标最高值，这表明此时生成的主题在语义连贯性与类别可分性方面达到最优，因此我们在MixSNIPS数据集上选择12作为主题数，在MixATIS数据集上选择26作为主题数。

5.5.2 消融实验

为了验证不同模块对本文模型的影响，我们在MixATIS公开数据集上进行了消融实验。实验结果如表 2所示。

(1) **w/o BiLSTM**：移除意图感知的BiLSTM和槽位感知的BiLSTM，仅使用BERT进行编码。实验结果表明，意图识别准确率和槽位填充$F1$值均有所下降，验证了BiLSTM在建模局部上下文关系方面的有效性。

(2) **w/o Topic-Guided Intent Enhancement**：移除基于主题引导的意图表示增强网络，直接使用文本编码器的编码特征进行解码。实验结果显示，意图识别准确率大幅下降，验证了主题增强特征在提升意图判别能力中的重要作用。

(3) **w/o Topic-Guided Slot Enhancement**：移除基于主题引导的槽位表示增强网络，直接使用文本编码器的编码特征进行解码。实验结果表明，槽位填充$F1$值下降，显示主题增强特征对提升槽位填充性能的实际贡献。

(4) **Slot Fusion (Cross-Attention instead of Concat)**：将基于主题引导的槽位表示增强网络中的拼接操作（Concat）替换为交叉注意力，以探索不同融合机制对模型性能的影响。实验发现，该替换导致槽位填充$F1$值下降，表明交叉注意力在槽位填充任务中对token表示产生重构性扰动，削弱了模型对局部边界的建模能力，从而验证了本文提出的基于主题引导的槽位表示增强网络设计的合理性。

(5) **Intent Fusion (Concat instead of Cross-Attention)**：将意图增强模块中的交叉注意力机制替换为拼接方式。实验结果显示，该替换导致意图识别性能下降，表明交叉注意力在刻画主题增强特征与意图特征间的深层语义关联中具有关键作用，能实现语义信息的动态对齐与选择性融合，相较于静态拼接更有效提升意图识别准确性。

5.5.3 案例分析

为进一步验证主题信息在意图识别与槽位填充任务中的有效性，我们对比分析了模型在保留完整结构与整体移除主题相关模块（即：主题提取模块、基于主题引导的意图表示增强网络和基于主题引导的槽位表示增强网络）的预测结果，通过具体案例分析主题信息对意图识别和槽位填充任务的具体贡献。结果如图 3所示。

(c) 2025 中国中文信息学会计算语言学专业委员会

Model	MixATIS			MixSNIPS		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
TA-IDSF	**88.9**	**81.9**	**50.9**	**96.9**	**97.5**	**84.8**
w/o BiLSTM	88.4	81.2	50.2	96.4	96.9	84.2
w/o Topic-Guided Intent Enhancement	88.7	78.6	49.1	96.6	94.3	83.1
w/o Topic-Guided Slot Enhancement	88.6	81.6	49.7	96.5	97.1	83.5
Slot Fusion(Cross-Attention instead of Concat)	88.1	81.5	50.3	95.8	97.3	84.0
Intent Fusion(Concat instead of Cross-Attention)	88.8	79.8	49.5	96.7	95.8	83.7

Table 2: TA-IDSF与现有模型在MixATIS 和MixSNIPS 数据集上的消融实验结果

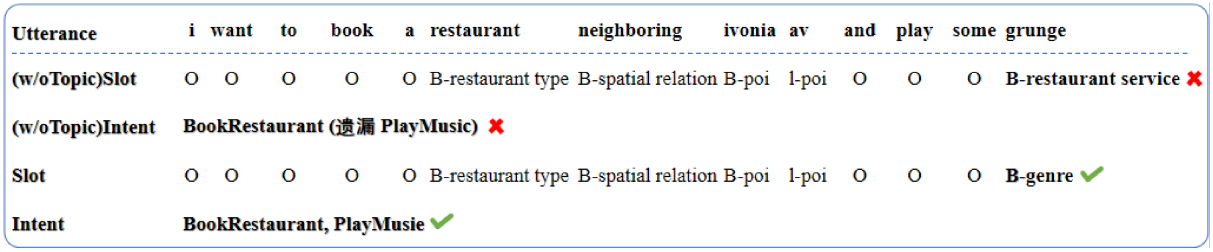


Figure 3: TA-IDSF的案例分析

从图 3可见，去除主题模块后，模型仅识别出前半句的“BookRestaurant”（预定餐厅）意图，未能检测到后半句的“PlayMusic”（播放音乐）意图。这是因为模型缺乏对语句主题倾向的理解，无法捕捉意图边界信号，倾向于单一意图。相比之下，完整模型正确预测了全部意图，主题模块通过文档-主题分布矩阵提供全局语义信息，提取出三个相关主题：主题4（餐厅预订：restaurant, book, cuisine）、主题10（音乐播放：play, genre, artist）和主题7（多任务连接：and, some）。主题4和10的高概率分布引导模型分割出“BookRestaurant”（预定餐厅）和“PlayMusic”（播放音乐）两个意图，主题7的and强化了逻辑连接词的分割作用。

针对槽位填充结果，后半句的“grunge”(垃圾摇滚)被错误标注为“B-restaurant_service”(餐厅服务)，而真实标签应为“B-genre”(艺术类型)。这是因为移除主题相关模块后，模型仅依赖文本编码器进行语义建模，缺乏对全局主题语义的感知，导致模型受前半句餐厅领域的词汇影响，误将“grunge”(垃圾摇滚)关联到餐厅服务，对后半句产生误导，引发槽位填充中的误差传播。而本文的模型通过引入的主题信息可感知句子中语义的潜在转变，根据主题10中包含的关键词“genre”(艺术类型)强化了对“grunge”(垃圾摇滚)所指向主题类别的建模，从而引导模型将其正确识别为B-genre(艺术类型)，显著缓解了因语义干扰而导致的槽位识别错误。

6 总结

本文提出了一种主题感知的意图识别与槽位填充联合建模方法(TA-IDSF)，TA-IDSF通过主题提取模块建模句子整体语义结构与主题分布倾向，从全局语义层面对两个子任务的建模过程进行引导，缓解了多意图场景中因全局语义建模不足引发的误差传播问题。通过本文的大量实验验证，可以得到以下结论：(1)引入全局语义信息可以有效提升意图识别与槽位填充两个任务的语义建模能力。(2)差异化的主题引导机制有助于根据任务需求平衡全局语义与局部特征，提升模型对不同任务的建模能力与适应性。

参考文献

Fengyu Cai, Wanhao Zhou, Fei Mi, and Boi Faltings. 2022. Slim: Explicit slot-intent mapping with bert for joint multi-intent detection and slot filling. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7607–7611. IEEE.

Lisong Chen, Peilin Zhou, and Yuexian Zou. 2022. Joint multiple intent detection and slot filling via self-distillation. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7612–7616. IEEE.

Yahui Chen. 2015. Convolutional neural network for sentence classification. Master's thesis, University of Waterloo.

(c) 2025 中国中文信息学会计算语言学专业委员会

Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, et al. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190*.

Rashmi Gangadharaiah. 2019. Joint multiple intent detection and slot labeling for goal-oriented dialog.

Alexander Genkin, David D Lewis, and David Madigan. 2007. Large-scale bayesian logistic regression for text categorization. *technometrics*, 49(3):291–304.

Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen. 2018. Slot-gated modeling for joint slot filling and intent prediction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 753–757.

Charles T Hemphill, John J Godfrey, and George R Doddington. 1990. 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*.

Vasima Khan and Tariq Azfar Meenai. 2021. Pretrained natural language processing model for intent recognition (bert-ir). *Human-Centric Intelligent Systems*, 1(5):66–74.

Hyun Kim, Jong-Hyeok Lee, and Seung-Hoon Na. 2017. Predictor-estimator using multilevel task learning with stack propagation for neural quality estimation. In *Proceedings of the second conference on machine translation*, pages 562–568.

Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. 2019. Text classification algorithms: A survey. *Information*, 10(4):150.

Gakuto Kurata, Bing Xiang, Bowen Zhou, and Mo Yu. 2016. Leveraging sentence-level information with encoder lstm for semantic slot filling. *arXiv preprint arXiv:1601.01530*.

Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

William Marslen-Wilson and Lorraine Komisarjevsky Tyler. 1980. The temporal structure of spoken language understanding. *Cognition*, 8(1):1–71.

Grégoire Mesnil, Xiaodong He, Li Deng, and Yoshua Bengio. 2013. Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding. In *Interspeech*, pages 3771–3775.

Nam Nguyen and Yunsong Guo. 2007. Comparisons of sequence labeling algorithms and extensions. In *Proceedings of the 24th international conference on Machine learning*, pages 681–688.

Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen, and Ting Liu. 2019. A stack-propagation framework with token-level intent detection for spoken language understanding. *arXiv preprint arXiv:1909.02188*.

Libo Qin, Xiao Xu, Wanxiang Che, and Ting Liu. 2020. Agif: An adaptive graph-interactive framework for joint multiple intent detection and slot filling. *arXiv preprint arXiv:2004.10087*.

Libo Qin, Fuxuan Wei, Tianbao Xie, Xiao Xu, Wanxiang Che, and Ting Liu. 2021. Gl-gin: Fast and accurate non-autoregressive model for joint multiple intent detection and slot filling. *arXiv preprint arXiv:2106.01925*.

Nguyen Anh Tu, Hoang Thi Thu Uyen, Tu Minh Phuong, and Ngo Xuan Bach. 2023. Joint multiple intent detection and slot filling with supervised contrastive learning and self-distillation. In *ECAI 2023*, pages 2370–2377. IOS Press.

Gokhan Tur and Renato De Mori. 2011. *Spoken language understanding: Systems for extracting semantic information from speech*. John Wiley & Sons.

Ngoc Thang Vu, Pankaj Gupta, Heike Adel, and Hinrich Schütze. 2016. Bi-directional recurrent neural network with ranking loss for spoken language understanding. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6060–6064. IEEE.

CCL 2025

(c) 2025 中国中文信息学会计算语言学专业委员会

Xiaobao Wu, Thong Nguyen, Delvin Zhang, William Yang Wang, and Anh Tuan Luu. 2024. Fastopic: Pretrained transformer is a fast, adaptive, stable, and transferable topic model. *Advances in Neural Information Processing Systems*, 37:84447–84481.

Puyang Xu and Ruhi Sarikaya. 2013. Convolutional neural network based triangular crf for joint intent detection and slot filling. In *2013 ieee workshop on automatic speech recognition and understanding*, pages 78–83. IEEE.

Shangjian Yin, Peijie Huang, Dongzhu Liang, Zhuoqi He, Qianer Li, and Yuhong Xu. 2023. 基于多意图融合框架的联合意图识别和槽填充. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics*, pages 54–63.

Steve Young, Milica Gašić, Blaise Thomson, and Jason D Williams. 2013. Pomdp-based statistical spoken dialog systems: A review. *Proceedings of the IEEE*, 101(5):1160–1179.

Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*.

Zhihong Zhu, Xuxin Cheng, Zhanpeng Chen, Zhichang Wang, Zhiqi Huang, and Yuexian Zou. 2024. Sala: Scenario-aware label graph interaction for multi-intent spoken language understanding. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 3570–3580.

徐梓翔, 车万翔, and 刘挺. 2017. 基于bi-lstm-crf 网络的语义槽识别. 智能计算机与应用, 7(6):91–94.

CCL 2025

(c) 2025 中国中文信息学会计算语言学专业委员会