

# 面向中文口语理解的基于依赖引导的字特征槽填充模型

朱展标, 黄沛杰\*, 张业兴, 刘树东, 张华林, 黄均曜

华南农业大学, 数学与信息学院, 广东广州, 510642

zhuzhanbiao@stu.scau.edu.cn, pjhuang@scau.edu.cn, yip\_hing@stu.scau.edu.cn,  
sudan@stu.scau.edu.cn, vol\_chris@scau.edu.cn, wishyao@stu.scau.edu.cn

## 摘要

意图识别和槽信息填充的联合模型将口语理解技术(Spoken language understanding, SLU)提升到了一个新的水平, 但由于存在出现频率低或未见过的槽指称项(0-shot slot mentions), 模型的序列标注性能受限, 而且这些联合模型往往没有利用输入序列存在的语法知识信息。已有研究表明序列标注任务可以通过引入依赖树结构, 辅助推断序列标注中槽的存在。在中文口语对话理解中, 由于中文话语是一串字序列组成, 输入话语的字和槽信息是一一对应的, 因而槽信息填充模型往往是字特征模型。基于词的依赖树结构无法直接应用于基于字特征的槽填充模型。为了解决字词之间的矛盾, 本文提出了一种基于字模型的依赖引导槽填充模型(Dependency-guided character-based slot filling model, DCSF), 提供了一种简洁的方法解决将词级依赖树结构引入中文字特征模型的冲突, 同时通过对话语中词汇内部关系进行建模, 保留了词级上下文信息和分词信息。在公共基准语料库SMP-ECDT和CrossWOZ上的实验结果表明, 我们的模型优于比较模型, 特别是在未见过的槽指称项和低资源情况下有很大的改进。

**关键词:** 口语对话理解; 槽信息填充; 依赖结构; 字特征模型

## A Dependency-Guided Character-Based Slot Filling Model for Chinese Spoken Language Understanding

Zhanbiao Zhu, Peijie Huang\*, Yexing Zhang, Shudong Liu,  
Hualin Zhang, Junyao Huang

College of Mathematics and Informatics, South China Agricultural University, China  
zhuzhanbiao@stu.scau.edu.cn, pjhuang@scau.edu.cn, yip\_hing@stu.scau.edu.cn,  
sudan@stu.scau.edu.cn, vol\_chris@scau.edu.cn, wishyao@stu.scau.edu.cn

## Abstract

The joint models for intent detection and slot tagging have taken the state of the art of spoken language understanding (SLU) to a new level. However, the presence of rarely seen or unseen mention degrades the performance of the model. Earlier research showed that sequence labeling task can benefit from the use of dependency tree structure for inferring existence of slot tags. In Chinese spoken language understanding, common models for slot filling are character-based hence word-level dependency tree structure can not be integrated into model directly. In this paper, we propose a dependency-guided character-based slot filling (DCSF) model, which provides a concise way to resolve the conflict of incorporating the word-level dependency tree structure into the

\*通讯作者

©2021 中国计算语言学大会

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

character-level model in Chinese. Our DCSF model can integrate dependency tree information into the character-level model while preserving word-level context and segmentation information by modeling different types of relationships between Chinese characters in the utterance. Experimental results on the public benchmark corpus SMP-ECDT and CrossWOZ show our model outperforms the compared models and has a great improvement, especially in low resource and unseen slot mentions scenario.

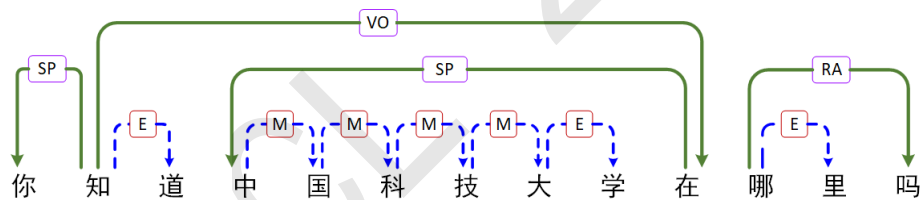
**Keywords:** Spoken language understanding , Slot filling , Dependency structure , Character-based model

## 1 引言

口语对话理解 (Spoken language understanding, SLU) 应用系统在我们的日常生活中变得越来越重要, 许多便携式设备提供的个人智能助理 (Personal Intelligent Assistant, PIA) 比如Amazon Alexa、Microsoft Cortana以及Apple Siri, 它们能够理解用户的指令, 帮助用户执行各种任务或与用户进行闲谈交流。这些智能助手的核心部分是口语对话理解, 其中口语对话理解旨在从用户的话语中形成语义框架, 其中包括意图检测 (Intent detection)和槽填充(Slot filling)两个部分 (Tür and Mori, 2011), 具体地说, 它能区分用户话语中的意图信息, 捕捉用户话语中的语义组成成分(Liu and Lane, 2016)。如图1.(a) 所示, 给定一个查询地图的话语: “你知道中国科技大学在哪里吗”, 话语中每个字与槽标签一一对应, 并且包含一个整个话语的意图。

话语	你	知	道	中	国	科	技	大	学	在	哪	里	吗
槽标签	0	0	0	B-poi	I-poi	I-poi	I-poi	I-poi	I-poi	0	0	0	0
意图	Query												

(a) 话语示例



(b) 话语中的依赖关系和词汇关系

Figure 1: 一个带有意图和槽信息标注(BIO 格式)的示例。话语存在查询(Query) 意图和景点 (Position of interest, poi) 的槽标签。话语中的依赖关系如绿色实线所示, 词汇关系如蓝色虚线所示

在之前的研究中, 意图检测任务和槽填充任务是分开训练的。意图检测任务通常被看作一个话语语义分类问题, 常见的意图识别方法分类方法包括支持向量机 (Support vector machine, SVM)和循环神经网络 (Recurrent neural network, RNN)。而槽信息填充可以被视为序列标注任务, 常见解决序列标注问题的方法有条件随机场(Conditional random fields, CRF)、最大熵马尔可夫模型(Maximum entropy Markov model, MEMMs)和RNN模型 (Yao et al., 2014)。考虑到槽填充与意图检测之间的强交互关系, 越来越多的研究将意图检测与槽填充任务结合起来, 一些研究者采用基于多任务框架的联合意图识别和槽填充模型 (Liu and Lane, 2016; Li et al., 2018; Goo et al., 2018; E et al., 2019; Qin et al., 2019)。一些研究者使用联合学习方法只考虑表面层面上的共享参数, 不能够直观地查看两者直接的交互关系, 缺少解释性(Liu and Lane, 2016; Hakkani-Tür et al., 2016; Chen et al., 2019)。一些模型提出将意图标签通过门控机制显性地传入到槽填充模型, 这种显性联合方式能够提高模型槽标签预测的表现并且具有一定解释性 (Goo et al., 2018; Qin et al., 2019)。也有研究者将槽标签和意图, 标签互相迭代交互传入到

模型中, 以此逐步提高两个任务模型表现(E et al., 2019)。此外, 还有一些研究者提出了一种基于动态路由算法的胶囊网络, 用于联合意图检测和槽信息填充任务(Zhang et al., 2019)。

当给予模型足够的训练数据的情况下, 上述模型可以取得很好的效果, 但是当出现频率低的槽指称项(Rare seen slot mentions, 或称为Few-shot slot mentions)或未见过的槽指称项(Unseen slot mentions, 或称为Zero-shot slot mentions)的时候, 这些模型的槽标签预测准确性会受到损失。特别是在预测长度较长的槽标签, 或者在低资源的场景中, 例如“中国科技大学”这样的长度较长且未在训练集出现过的槽指称项往往不能够被正确识别出对应的槽标签, 因而模型的性能受到影响。

此外, 这些上下文感知模型只关注于应用上下文语义来识别槽信息, 往往没有利用外部知识, 例如句子中的语法知识。许多早期的研究表明, 序列标记任务如命名实体识别(Name entity recognition, NER)任务, 模型能从使用依赖树结构的受益(Huang et al., 2017; Jie and Lu, 2019; Nie et al., 2020)。一方面, 依赖树结构可以捕捉单词之间的长距离依赖关系; 另外一方面, 依赖结构可以捕获单词和实体之间的语义级交互, 它可以潜在地推断槽标签或实体的边界, 特别是长度较长的槽位。这些依赖结构信息能够帮助模型正确识别槽边界(Jie and Lu, 2019)。

中文的口语对话理解相较于英文口语对话理解更具挑战, 因为口语话语理解的中文话语由一串字序列组成的话语, 是未进行预先分词的。已有研究表明基于中文字特征的序列标注模型可以取得比完全基于词特征的序列标注模型更好性能(Zhang and Yang, 2018)。但是汉语的依赖结构是基于词级别的, 无法直接集成到字级特征模型中, 如汉语的BERT (Bidirectional Encoder Representations from Transformers)模型(Devlin et al., 2019)中。针对这一问题, 本文提出了一种依赖引导的基于字特征的槽填充模型(DCSF), 该模型能够解决字级口语对话模型与词级依赖树结构之间的冲突。我们的DCSF模型提供了一个简洁的方式, 建模输入层两种不同类型关系, 其中包含依赖关系(图1中的绿色弧线(b))和词内部关系(图1中的虚线蓝弧(b))。考虑到不同类型的关系, 我们的模型能够一方面捕获基于依赖的远距离交互, 可以捕获单词和实体之间的语义级交互; 另一方面由于建立词内部联系保留了词级信息, 增强模型输入字表征的表达能力和区分槽边界的能力。

本文主要贡献如下:

- (1) 提出了一种基于依赖引导的字特征槽填充模型, 该模型解决了字模型和词级依赖树结构之间的冲突, 有效地实现了在字级特征的中文口语话语槽填充模型中融入词级的依赖关系。
- (2) 在公开基准数据集上的实验结果表明, 本文的模型在槽位提取F1和话语准确率两方面均取得了显著优于目前研究进展口语理解模型的性能。消融实验、0-shot槽位的提取和低资源实验的定量分析进一步证明了本文提出的模型的良好效果。

## 2 相关工作

本文的研究是将依赖结构引入到采用字特征的中文口语理解的槽填充模型中, 下面我们简要介绍一下口语理解和基于依赖结构的序列标注两方面的相关工作。

### 2.1 口语理解

口语理解(spoken language understanding, SLU)目的是抽取用户话语的语义框架, 通常涉及到用户话语的意图检测和槽填充两个任务(Tür and Mori, 2011; 俞凯等, 2015; 赵阳洋等, 2020)。其中意图检测被看成是分类任务, 常用方法包括SVM(Haffner et al., 2003)、Adaboost(Schapire and Singer, 2000)。而槽填充任务被看成是序列标注任务, 传统的方法是使用具有较强序列标注能力的CRF结构(Raymond and Ricciardi, 2007)。近年来, 由于深度学习强大的学习能力, 越来越多的模型使用深度学习框架去提升意图检测效果, 例如卷积神经网络(convolutional neural network, CNN)(Yoon Kim, 2014)、RNN(Mesnil et al., 2015)等方法。在槽信息填充方面, Yao等人(2014)提出基于LSTM模型的槽填充方法, 缓解了循环神经网络记忆遗忘问题。研究者还进一步使用CRF作为句子级的优化标准, 使得模型能够处理输出序列之间的依赖关系(Yao et al., 2014)。

考虑到分开处理意图检测和槽填充任务, 模型不能够充分利用两者之间的交互的信息。因此, Zhang和Wang(2016)提出了使用循环神经网络联合意图检测和槽填充模型来解决上

述问题。而Liu和Lane (2016)则将注意力机制引入到了SLU的联合学习模型中。随后, Li等人 (2018)、Goo等人 (2018)、E等人 (2019)、Qin等人 (2019)尝试在SLU的联合学习模型中显式建立意图检测和槽填充的联系。最近, 研究者们将上下文相关的单词表征 (如BERT) 应用到SLU任务中 (Chen et al., 2019), 进一步提高了SLU的性能。尽管SLU任务已经取得了较大的进展, 但在少样本槽位提取方面仍然存在一些不足。已有的研究表明序列标记任务可以通过引入依赖树结构, 辅助推断标签的存在, 提高序列标记的准确度 (Jie and Lu, 2019)。因此, 本文在研究进展的SLU模型的基础上, 研究依赖引导的槽填充模型。

### 2.2 基于依赖结构的序列标注

依赖树结构可以捕捉单词之间的长距离依赖关系, 捕获词语和实体之间的语义级联系。研究者们发现引入依赖结构能够提升模型序列标注能力。Ling和Weld (2012)引入实体在依赖树结构中的依赖词语作为特征输入, 提升了模型实体识别的能力。Liu等人 (2010)提出使用Skip-chain CRF链接具有不同关联关系的词语以提升模型表现。Jie等人 (2019)提出依赖树引导模型, 应用于NER任务。然而, 由于现有的中文依赖树结构是基于词级的, 而研究进展的中文序列标注模型采用的是字特征的模型 (Zhang and Yang, 2018), 这使得依赖结构难以直接应用到中文槽填充任务。为此, 本文对中文话语中字之间不同类型的关系进行建模, 解决基于中文字特征的模型与基于词的依赖树结构在应用结合时的矛盾。

## 3 方法

在本节中, 将介绍用于本文提出的用于SLU任务的依赖引导的基于字特征槽填充(DCSF)模型。图2显示了该模型示例话语“查询天龙集团的股票”及其依赖关系结构。模型的结构包括字特征嵌入层、关系嵌入层、双向LSTM层和输出预测层。

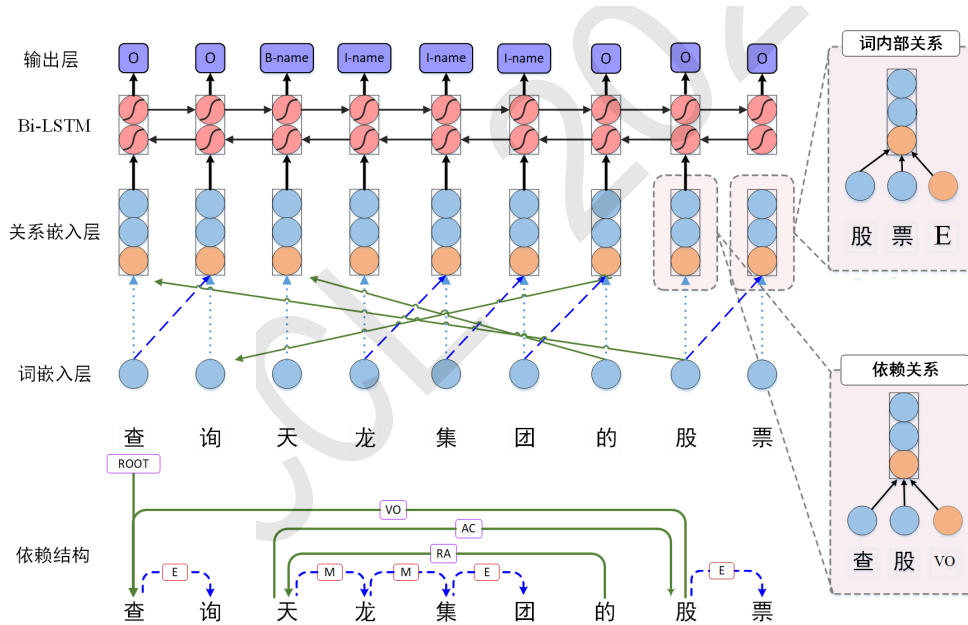


Figure 2: DCSF模型的架构 (本文的重点是改进SLU任务中的槽信息填充任务, 因此只展示基于BERT的SLU联合学习模型中槽填充的部分)

### 3.1 输入表征

为了捕获远距离交互, 我们在模型中应用了依赖信息。在依赖结构中, 一个话语中的每个词都只有一个父节点(词根除外)。我们将词的依赖结构之间的关系表示为依赖关系, 我们将相邻字之间的关系表示为词内部关系, 提供了内部联系和词边界信息。关于这依赖关系和词内部关系的详细信息可以在图2的子图中看到。

**依赖关系** 由于汉语依赖树分析是基于词的, 如果采用完全词级别模型, 会出现边界错误, 进而将错误传播到后续槽信息填充预测任务中。为了使该过程与字模型兼容并且缓解分词错误

带来的不良影响，在基于字特征的模型中，我们只选择词的第一个字来表示依赖关系。对于依赖关系， $(x_p, x_t, r)$ 以 $x_p$ 为父结点， $x_t$ 为子结点， $r_{dep}(r_{dep} \in \{ROOT, VO, \dots, SP\})$ 为依赖关系。如图2中的示例说明，如父结点词 $x_p$ “天龙集团”和子结点词 $x_t$ “股票”存在依赖关系，为了与字模型兼容并且缓解分词错误带来的不良影响，父结点词 $x_p$ “天龙集团”和子结点词 $x_t$ “股票”的依赖关系将转换成单个字 $c_t$ “天”和 $c_p$ “股”的关系。受Jie等人(2019)的启发，我们将子结点的字特征表示与其父节点字表示及其对应依赖关系表示拼接在一起。

$$c_t = First(x_t) \quad (1)$$

$$e_t = Embed_c(c_t) \quad (2)$$

$$v_{dep} = Embed_r(r_{dep}), (r_{dep} \in \{ROOT, VO, \dots, SP\}) \quad (3)$$

$$u_{dep}^{(t)} = [e_t; e_p; v_{dep}], x_p = parent(x_t) \quad (4)$$

其中 $e_t$ 和 $e_p$ 分别是词 $x_t$ 和词 $x_p$ 中第一个字的字特征向量， $v_{dep}$ 是依赖关系 $r_{dep}$ 的关系标签的表示向量。 $Embed_c$ 和 $Embed_r$ 分别是字嵌入层和关系标签嵌入层。细节可以在图2中的依赖关系子图中看到。

**词内部关系** 目前常用的汉语序列标注方法大多将单词信息整合到基于字特征的模型中，但大多只考虑了静态拼接词嵌入，缺乏灵活性(Zhang and Yang, 2018; Liu et al., 2019)。此外，边界信息在序列标注任务，特别是槽填充任务中也有应用价值，仅仅加入静态的词嵌入往往丢失了这部分边界信息。为了利用动态词嵌入信息，我们将一个词中的连续字用一个特殊的弧连接起来，包括M(单词中间)和E(单词结尾)。我们将该字特征向量与其前向字特征向量和对应关系嵌入向量拼接在一起。

$$v_{word} = Embed_r(r_{word}), (r_{word} \in \{M, E\}) \quad (5)$$

$$u_{word}^{(t)} = [e_t; e_f; v_{word}], c_f = forward(c_t) \quad (6)$$

其中， $e_t$ 和 $e_f$ 分别是词组中字 $c_t$ 及其相邻前向字 $c_f$ 的字表示。 $v_{word}$ 是词内部关系 $r_{word}(r_{word} \in \{M, E\})$ 中的嵌入表示向量。例如词“股票”，它包含两个字分别是“股”和“票”。“票”前一个字是“股”，为了表示这种词内部关系，我们将“股”的字特征向量和“票”的字特征向量和词内部关系嵌入向量E(E代表“票”是词语“股票”最后一个字)拼接在一起。细节可以在图2中的词内部关系子图中看到。

### 3.2 模型结构

在图2中的话语例子“查询天龙集团的股票”展示了输入序列之间的关系，图中可以看出依赖关系和单词关系可以提供不同的信息。例如，“查询”(搜索)和“股票”(股票)是分开的话语，但是在语法树却是邻居。通过引入远距离依赖关系，模型可以捕获远距离依赖信息。此外，模型还考虑了词内部关系，能够捕获词的上下文信息和词的边界信息。每个位置的隐藏状态输入到Bi-LSTM层后，会传播到下一个预测层，用于标记槽位标签。图2所示的架构可以有效地嵌入远距离的父依赖字，以连接父结点信息和依赖关系。因而该模型能够间接捕获长距离交互作用、动态词信息和词边界信息。知识增强的序列向量为 $\{u^{(1)}, u^{(2)}, \dots, u^{(T)}\}$ ,  $u \in \{u_{dep}, u_{word}\}$ , 其中 $T$ 是句子长度。将知识增强后的序列向量输入到Bi-LSTM (Hochreiter and Schmidhuber, 1997; Yao et al., 2014)中，其中Bi-LSTM模型内部结构如下定义：

$$\vec{h}_t = LSTM(u_t, \vec{h}_{t-1}) \quad (7)$$

$$\overleftarrow{h}_t = LSTM(u_t, \overleftarrow{h}_{t-1}) \quad (8)$$

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (9)$$

其中 $t$ 代表这个第 $t$ 个时间步输入，最终将Bi-LSTM的隐藏层向量 $h_t$ 输入到槽信息填充预测层中：

$$y_t^{slot} = \text{softmax}(W_h^{slot} h_t) \quad (10)$$

$$o_i^{slot} = \text{argmax}(y^{slot}) \quad (11)$$

其中 $W^{slot}$ 是槽标签预测的权重矩阵， $o_i^{slot}$ 最终预测的槽标签。对于意图识别，Qin等人(2019)发现使用字级别意图能够达到更好的模型表现效果，与槽信息填充预测过程相似，我们将Bi-LSTM隐藏层输入到意图预测层中：

$$y_t^{intent} = \text{softmax}(W_h^{intent} h_t) \quad (12)$$

$$o_t^{intent} = \text{argmax}(y^{intent}) \quad (13)$$

其中 $W^{intent}$ 是槽标签预测的权重矩阵， $o_i^{intent}$ 最终预测的槽标签。意图识别和槽信息填充的损失函数定义是：

$$\mathcal{L}_{intent} \triangleq - \sum_{t=1}^T \sum_{j=1}^{n^{intent}} \hat{y}_j^{t,intent} \log(y_j^{t,intent}) \quad (14)$$

$$\mathcal{L}_{slot} \triangleq - \sum_{t=1}^T \sum_{j=1}^{n^{slot}} \hat{y}_j^{t,slot} \log(y_j^{t,slot}) \quad (15)$$

其中 $t$ 代表的是话语中第 $t$ 个字， $y_j^{t,slot}$ 和 $y_j^{t,intent}$ 是真实槽标签和真实意图标签， $n^{intent}$ 和 $n^{slot}$ 分别代表着意图标签的个数和槽标签的个数。为了能够联合两个模型，我们将两个任务损失函数联合，总的损失函数为：

$$\mathcal{L} = \mathcal{L}_{intent} + \mathcal{L}_{slot} \quad (16)$$

## 4 实验

### 4.1 数据集

我们在SMP-ECDT 2019<sup>1</sup>和CrossWOZ (Zhu et al., 2020)的公共基准文本语料上对我们的模型进行训练和评估。SMP-ECDT 2019 (Social Media Processing - the Evaluation of Chinese Human-Computer Dialogue Technology)是来自ASR系统的转录。SMP数据集有24种不同的意图类型和124种不同的槽位标签。CrossWOZ是大规模的中文跨域Wizard-of-Oz任务导向数据集，分别包含5个领域和72种不同类型的槽位。SMP和CrossWOZ数据集的统计结果见表1。在SMP和CrossWOZ中测试数据集0-shot的槽指称项比例分别为70.31%和26.26%。因为SMP数据集的训练数据量较小，并且在测试集中存在大量的0-shot槽，所以SMP数据集更具有挑战性。

### 4.2 超参数

在实验中，我们使用HaNLP<sup>3</sup>工具来捕获依赖结构，其性能优于Stanford CoreNLP toolkit<sup>4</sup> (Manning et al., 2014)。为了避免过拟合，Dropout (Srivastava et al., 2014)被设置为0.1到0.5进行验证，实验中使用了Adam优化器 (Kingma and Ba, 2015)，学习速率为1e-5，批量大小 (Batch size) 设置为32，L2正则化参数为1e-8，关系嵌入维度为50。此外，我们使用交叉熵作为意图检测和槽填充任务的损失函数。在没有使用预训练字向量模型中，我们均采用随机初始化学向量的方式。以下结果所示数据均为10个独立实验的平均值。

<sup>1</sup><http://conference.cipsc.org.cn/smp2019/evaluation.html>

<sup>2</sup><https://github.com/OnionWang/SMP2019-ECDT-NLU>

<sup>3</sup><https://github.com/hankcs/HanLP/>

<sup>4</sup><https://stanfordnlp.github.io>

	SMP	CrossWOZ
训练集数量	1656	14962
验证集数量	414	1518
测试集数量	509	1590
槽标签数量	124	72
槽指称项数量	485	457
0-shot槽指称项数量	341	120
0-shot槽指称项比例	70.31%	26.26%

Table 1: SMP 和CrossWOZ 数据集的统计结果

### 4.3 对比的方法

我们将DCSF模型与以下方法进行了比较，包括：

- Attention BiRNN. Liu和Lane (2016)利用注意力机制，让模型网络学习信息槽和意图之间的关系。
- Self-Attentive Model. Li等人 (2018)提出了一种新的具有意图增强门控机制的自注意力模型，以利用信息槽与意图之间的语义相关性。
- Slot-Gated Atten. Goo等人 (2018)提出了信息槽门控联合模型，以更好地探究槽填充与意图检测的相关性。
- SF-ID Network. E等人 (2019)介绍了一种SF-ID网络，为槽填充和意图检测建立起直接性的连接，以帮助它们相互促进与提升。
- Stack-propagation. Qin等人 (2019)采用stack-propagation框架，提出了一个结合单词级别意图识别的联合学习模型。
- Joint BERT. Chen等人 (2019)应用BERT模型进行联合槽填充和意图检测。

以上方法中，DCSF与Attention BiRNN、Self-Attentive Model、Slot-Gated Atten、SF-ID Network、Stack-propagation均是没有采用预训练字向量的模型。DCSF(BERT)和Joint BERT采用了预训练模型BERT。

### 4.4 主实验

与已有的工作 (Goo et al., 2018; E et al., 2019; Chen et al., 2019)一样，我们在槽填充中使用F1值作为评价指标，意图预测中使用正确率作为评价指标，句子级正确率使用整句正确率来综合评估口语理解的性能，表2展示了我们提出的DCSF模型的实验结果，并在SMP-ECDT数据集和CrossWOZ数据集上进行了比较与分析。由于本文的重点是改进口语理解中的信息槽填充，因此我们主要观察不同模型在槽填充F1值和句子整句正确率上的表现。

模型	SMP			CrossWOZ		
	意图识别	槽位提取	整句	意图识别	槽位提取	整句
Attention BiRNN (Liu and Lane, 2016)	85.36	54.91	49.85	95.01	87.20	77.63
Self-Attentive Model (Li et al., 2018)	85.51	51.41	45.42	95.09	85.27	74.71
Slot-Gated Full Atten (Goo et al., 2018)	84.81	60.88	54.21	95.15	91.03	81.13
SF-ID Network (E et al., 2019)	86.51	64.52	56.33	95.84	91.20	82.64
Stack-propagation (Qin et al., 2019)	88.65	71.62	63.78	95.85	90.57	81.77
DCSF	89.14	73.09	65.52	95.53	91.72	82.95
Joint BERT (Chen et al., 2019)	92.46	82.31	75.42	95.83	92.61	84.28
DCSF(BERT)	<b>93.00</b>	<b>84.31</b>	<b>77.23</b>	<b>95.94</b>	<b>92.79</b>	<b>84.57</b>

Table 2: 我们的模型和对比方法在测试集上的性能对比

从表2中可以看出, 在没有采用预训练字向量下, 我们提出的DCSF在两个公共基准语料库上的所有指标方面显著优于其他所有模型。在使用预训练模型条件下, 与Joint BERT的模型相比, DCSF (BERT) 模型在槽位提取(F1)和整句正确率得分上有较大的提升, 在意图(Acc)得分上略有提高。这些结果证明了远距离依赖信息和词边界信息的有效性, 而且这种远距离依赖信息和词边界信息不仅适用于没有采用预训练字向量的模型, 也适用于基于预训练BERT的模型。

#### 4.5 进一步研究

通过比较以上结果可以看出DCSF模型取得了良好的性能, 我们想进一步探究模型能有所提升的原因。我们首先进行了消融实验, 以分析我们的模型建模的不同类型的关系对模型整体性能带来的影响。然后, 我们给出了一个关于不同模型的0-shot槽位提取正确率的对比分析。

**消融实验** 为了研究不同类型的关系在DCSF模型中的贡献, 我们对DCSF(BERT)进行了消融实验。我们结果如表3所示。

模型	关系	SMP			CrossWOZ		
		意图识别	槽位提取	整句	意图识别	槽位提取	整句
消去外部关系	{M, E}	93.15	83.82	76.89	95.82	92.64	84.38
消去词内部关系	{dep}	92.68	83.69	76.65	95.82	92.68	84.42
+ 词向量	{dep}	92.63	83.83	76.79	95.76	92.70	84.48
消去边界关系E	{dep, M}	92.73	83.71	76.67	95.87	92.51	84.23
+ 词向量	{dep, M}	91.90	83.66	76.11	95.81	92.64	84.31
完整模型	{dep, M, E}	93.00	84.31	77.23	95.94	92.79	84.57

Table 3: DCSF(BERT) 消融实验

结果表明, 我们的模型建模的不同关系对于模型的整体性能都是有价值的。详细的分析如下:

- 消去外部关系: 在这种情况下, 我们从模型中删除外部关系。从结果中, 我们可以看到在SMP数据集中槽位提取F1分数下降了0.49%, 整体句子正确率分数下降了0.34%, 在CrossWOZ数据集中槽位提取F1分数下降了0.15%, 整体句子正确率分数下降了0.19%。这表明, 在话语中对词语之间依赖关系的建模, 有助于提高槽填充的性能, 并提高句子准确率。
- 消去内部关系: 在这个设置中, 我们从模型中删除了内部关系X和E。结果显示, 槽位提取F1分数明显下降, 这表明捕捉内部关系是有效的且对槽填充的提升效果明显。此外, 我们用词向量代替我们的内部关系 (Zhang and Yang, 2018; Ma et al., 2020), 该方法也可以代表对整个词语的信息进行建模。我们可以看到, “+词向量”版本优于“消去内部关系”版本而低于我们的完整模型。这进一步证明了词语信息对模型提升的有效性以及我们的内部关系机制对词语信息的捕捉效果。
- 消去边界E关系: 在实验中, 我们也观察了在我们的内部关系中只删除E关系的模型和使用词向量代替E关系的模型。从结果中, 我们可以看到所有指标的分数都下降了。这表明, 在我们的内部关系机制中, E关系可以更好地模拟出词语的边界信息。

**0-shot 实验** 由于存在在训练集未出现的槽指称项(0-shot slot mentions), 基线模型可能无法预测这些0-shot槽指称项, 给模型表现带来了巨大的挑战, 尤其在SMP数据集中只包含1656条训练话语样本, 由于训练样本较小, 问题更加突出, 测试集包含的70.31%的槽指称项未在训练集样本出现过, 也即是0-shot槽指称项, 因此基线模型的性能受到严重损失。0-shot槽位提取和总体槽位提取召回率结果如表4所示。实验结果表明, 通过整合依赖关系和词内部关系, 我们提出的DCSF模型在使用普通词向量条件下, 0-shot槽位提取相较于对比模型取得了显著的提升, 而且由于SMP存在0-shot槽指称项问题更加突出, 因而DCSF模型在SMP上提高了更加明显。在基于BERT的方法模型中, DCSF在0-shot槽上也实现了较大的性能改进, 达到了



近2.39%，而在CrossWoz上也提高了2.13%。此外，在测试集中，基线模型0-shot上的表现比整体的召回率低，这表明话语中0-shot槽确实给槽填充带来挑战。根据实验结果表明，将依赖关系和词汇关系相结合，可以提高模型对槽指称项的识别能力，特别是在含有较多比例0-shot样本的语料中提升更加明显。

模型	SMP		CrossWOZ	
	0-shot	总体	0-shot	总体
Attention BiRNN (Liu and Lane, 2016)	48.64	58.89	63.43	91.50
Self-Attentive Model (Li et al., 2018)	38.56	55.07	71.66	94.78
Slot-Gated Full Atten (Goo et al., 2018)	44.05	64.86	53.38	84.77
SF-ID Network (E et al., 2019)	54.38	64.66	69.52	95.44
Stack-propagation (Qin et al., 2019)	62.73	71.57	80.67	93.35
DCSF	66.37	72.36	83.74	96.27
Joint BERT (Chen et al., 2019)	80.76	85.59	87.13	96.74
DCSF(BERT)	<b>83.15</b>	<b>86.84</b>	<b>89.26</b>	<b>97.02</b>

Table 4: DCSF模型和对比模型在0-shot槽位提取召回率

**低资源实验** 为了模拟低资源场景，我们对SMP和CrossWOZ的原始训练集进行了下采样，随机抽取了原始训练集数据的10%、25%、50%以及75%数据量，验证集和测试集数据不进行修改，和主实验保持一致。图3显示了在使用和不使用我们机制情况下，使用不同大小的训练数据，基于预训练BERT的模型在SMP和CrossWOZ两个公开数据集上的相对改进。实验结果表明，将依赖关系和词内部关系相结合引入模型中，确实有利于字特征模型的性能的进一步提升。特别在低资源场景下，由于仅给定少量的话语训练数据，传统的联合模型在低资源场景下不能充分学习到上下文语义，因而显式的依赖关系是必要的，因为它们可以提供语法信息和捕获长距离信息。此外，由于实体边界通常与某些单词边界重合，词内部关系可以帮助识别罕见和0-shot的槽指称项，这些词内部关系提供了丰富的边界信息。在仅有10%的训练语料的实验设定中，在SMP数据集和CrossWoz数据集，本文的DCSF(BERT)比Joint BERT的槽位提取F1值的提升分别达到7.07%和4.58%。

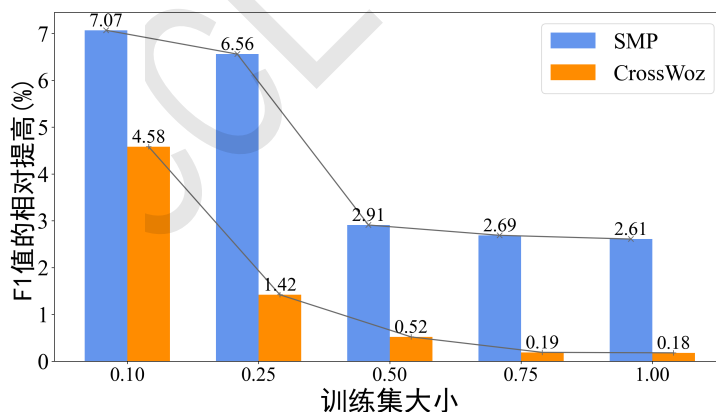


Figure 3: 在不同比例训练集下DCSF(BERT)相比Joint BERT的槽位提取F1值的相对提升

## 5 结论

本文将依赖树结构应用于口语对话理解模型中的基于字特征的联合意图识别和槽信息填充模型中，提出了一种依赖引导的基于字特征槽填充(DCSF)模型，解决了中文字特征模型和词级依赖树结构应用结合之间的矛盾。我们的DCSF模型以研究进展的基于BERT的联合学习模型为基础，提出了一种将词级别依赖信息融合到字级模型中方法，通过建模话语中存在的依赖

关系和词内部的连接关系，能够让模型捕捉到话语中长距离的依赖关系以及词的边界信息。在两个公共基准语料库SMP-ECDT和CrossWOZ上的实验结果表明，我们的模型在槽位(F1)和整句(Acc)得分上显著优于比较模型。消融研究证明了不同类型的关系对模型的整体性能都有贡献。对比不同模型在0-shot槽指称项和低资源情景的影响，进一步验证了所提模型的有效性。

## 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)的资助。

## 参考文献

- 俞凯, 陈露, 陈博等. 2015. 任务型人机对话系统中的认知技术——概念、进展及其未来. 计算机学报, 38 (12):2333-2348.
- 赵阳洋, 王振宇, 王佩, 杨添, 张睿, 尹凯. 2020. 任务型对话系统研究综述. 计算机学报, 43 (10): 1862-1896.
- Qian Chen, Zhu Zhuo, and Wen Wang. BERT for joint intent classification and slot filling. *CoRR*, abs/1902.10909, 2019.
- Jacob Devlin, Ming-Wei Chang, and Kenton Lee, et al. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, pages 4171-4186, 2019.
- Haihong E, Peiqing Niu, and Zhongfu Chen, et al. 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*, pages 5467-5471, 2019.
- Chih-Wen Goo, Guang Gao, and Yun-Kai Hsu, et al. 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, NAACL-HLT 2018*, pages 753-757, 2018.
- Patrick Haffner, Gökhan Tür, and Jerry H. Wright. Optimizing svms for complex call classification. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '03*, pages 632-635, 2003.
- Dilek Hakkani-Tür, Gökhan Tür, and Asli Çelikyılmaz, et al. Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association, INTERSPEECH 2016*, pages 715-719, 2016.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735-1780, 1997.
- Lifu Huang, Avirup Sil, and Heng Ji, et al. Improving slot filling performance with attentive neural networks on dependency structures. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017*, pages 2588-2597, 2017.
- Zhanming Jie and Wei Lu. Dependency-guided LSTM-CRF for named entity recognition. 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*, pages 3860-3870, 2019.
- Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*, pages 1746-1751, 2014.
- Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015*, 2015.
- Changliang Li, Liang Li, and Ji Qi. A self-attentive model with gate mechanism for spoken language understanding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, pages 3824-3833, 2018.

- Xiaonan Li, Hang Yan, Xipeng Qiu, and Xuanjing Huang. FLAT: chinese NER using flat-lattice transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020*, pages 6836–6842, 2020.
- Xiao Ling and Daniel S. Weld. Fine-grained entity recognition. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- Wei Liu, Tongge Xu, and QingHua Xu, et al. An encoding strategy based word-character LSTM for chinese NER. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, pages 2379–2389, 2019.
- 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*, pages 685–689, 2016.
- Jingchen Liu, Minlie Huang, and Xiaoyan Zhu. Recognizing biomedical named entities using skip-chain conditional random fields. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, BioNLP@ACL 2010*, pages 10–18, 2010.
- Ruotian Ma, Minlong Peng, and Qi Zhang, et al. Simplify the usage of lexicon in chinese NER. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020*, pages 5951–5960, 2020.
- Christopher D. Manning, Mihai Surdeanu, and John Bauer, et al. The stanford CoreNLP natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014*, pages 55–60, 2014.
- Grégoire Mesnil, Yann N. Dauphin, and Kaisheng Yao, et al. Using recurrent neural networks for slot filling in spoken language understanding. *IEEE ACM Trans. Audio Speech Lang. Process.*, 23(3):530–539, 2015.
- Yuyang Nie, Yuanhe Tian, Yan Song, Xiang Ao, and Xiang Wan. Improving named entity recognition with attentive ensemble of syntactic information. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, EMNLP 2020*, pages 4231–4245, 2020.
- Libo Qin, Wanxiang Che, and Yangming Li, et al. 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*, pages 2078–2087, 2019.
- Christian Raymond and Giuseppe Riccardi. Generative and discriminative algorithms for spoken language understanding. In *Proceedings of the 8th Annual Conference of the International Speech Communication Association, INTERSPEECH 2007*, pages 1605–1608. ISCA, 2007.
- Robert E. Schapire and Yoram Singer. Boostexter: A boosting-based system for text categorization. *Mach. Learn.*, 39(2/3):135–168, 2000.
- Nitish Srivastava, Geoffrey E. Hinton, and Alex Krizhevsky, et al. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, 2014.
- G. Tür and R. Mori. 2011. *Spoken language understanding: Systems for extracting semantic information from speech*, John Wiley and Sons, Inc.
- Ashish Vaswani, Noam Shazeer, and Niki Parmar, et al. Attention is all you need. In *Proceedings of the 2017 Annual Conference on Neural Information Processing Systems, NIPS 2017*, pages 5998–6008, 2017.
- 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*, pages 4077–4081, 2014.
- Kaisheng Yao, Baolin Peng, and Yu Zhang, et al. Spoken language understanding using long short-term memory neural networks. In *2014 IEEE Spoken Language Technology Workshop, SLT 2014*, pages 189–194, 2014.

- Xiaodong Zhang and Houfeng Wang. A joint model of intent determination and slot filling for spoken language understanding. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016*, pages 2993–2999, 2016.
- Chenwei Zhang, Yaliang Li, Nan Du, Wei Fan, and Philip S. Yu. Joint slot filling and intent detection via capsule neural networks. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019*, pages 5259–5267, 2019.
- Qi Zhu, Kaili Huang, Zheng Zhang, Xiaoyan Zhu, and Minlie Huang. CrossWOZ: A large-scale Chinese cross-domain task-oriented dialogue dataset. *Trans. Assoc. Comput. Linguistics*, 8:281–295, 2020.
- Yue Zhang and Jie Yang. Chinese NER using lattice LSTM. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018*, pages 1554–1564, 2018.

JCL 2021