

文章编号: 1003-0077(2023)02-0119-10

基于小样本学习的意图识别与槽位填充方法

孙相会¹, 苗德强¹, 窦辰晓¹, 袁龙², 马宝昌¹, 邓勇¹, 张露露¹, 李先刚¹

(1. 贝壳找房 人工智能技术中心, 北京 100085;

2. 南京理工大学 计算机科学与工程学院, 江苏 南京 210094)

摘要: “意图识别”与“槽位填充”是智能人机交互中的两个核心任务,受到学术界和工业界的广泛关注。目前业界前沿主流的方法,在一些学术公开数据集上已取得了不错的效果。不过这些方法大多依赖于丰富的标注数据集来完成训练,而数据集需要经过人工采集、标注等流程构造,且需满足其分布的均匀性。然而,真实业务场景下的数据却很难达到以上标准,往往面临小样本学习困难的难题,大多主流方法在小样本学习任务上的表现远不如其在大样本学习上的效果。针对此业界难点,该文提出一种基于半监督学习与迁移学习的“意图识别”与“槽位填充”的串联方法。该方法根据“意图识别”和“槽位填充”的各自任务特性,分别设计了针对性更强的小样本解决思路,即通过半监督学习的思想,在不需引入大量标注数据的情况下,利用无标签数据丰富、构造训练样本集,提高小样本意图识别的准确性;以及通过迁移学习的思想,将从大样本数据中学习到的先验知识迁移到小样本数据模型中,利用大样本数据与小样本数据间的公共知识,提高小样本槽位填充的精准度。该文所提出的方法通过实验对比被证实有效,且在 2021 年中国计算机学会大数据与计算智能大赛(CCF-BDCI)组委会与中国中文信息学会(CIPS)共同举办的全国信息检索挑战杯(CCIR Cup)的“智能人机交互自然语言理解”赛道取得了第一名的成绩。

关键词: 小样本学习; 半监督; 迁移学习

中图分类号: TP391

文献标识码: A

A Few-shot Learning Approach to Intent Recognition and Slot Filling

SUN Xianghui¹, MIAO Deqiang¹, DOU Chenxiao¹,

YUAN Long², MA Baochang¹, DENG Yong¹, ZHANG Lulu¹, LI Xiangang¹

(1. Center of Artificial Intelligence, KE Holdings Inc, Beijing 100085, China;

2. School of Computer Science and Engineering, Nanjing University of Science and Technology,
Nanjing, Jiangsu 210094, China)

Abstract: "Intent Recognition" and "Slot Filling" are two core tasks in intelligent human-computer interaction, which have received extensive attentions from academia and industry. Most state-of-the-art models on few-shot learning tasks are far inferior to their performance on many-shot learning tasks. In this paper, we propose a novel joint model based on semi-supervised and transfer learning for intent recognition and slot filling. Semi-supervised learning is used to identify the few-shot intent, requiring no additional labelled data. Transfer learning is used to exploit prior knowledge learned from the large samples to acquire a slot-filling model for the small samples. The proposed method won the first place in 2021 China Computer Society Big Data and Computational Intelligence Competition (CCF-BDCI) Organizing Committee and Chinese Information Society of China (CIPS) jointly held the National Information Retrieval Challenge Cup (CCIR Cup) track of 2021 CCIR Cup held by CCF Big Data & Computing Intelligence Contest (CCF-BDCI).

Keywords: few-shot learning; semi-supervised learning; transfer learning

0 引言

语音交互是任务型对话的核心模块,是智能人机交互领域中非常重要的研究方向,近年来受到学术界和工业界的广泛关注。用户仅需要通过与 AI 产品对话即可下达命令来控制操作 AI 设备,如播放音乐、开关电器等。为了能够准确解析用户的口令,通常我们需要对用户口令中的意图进行识别以及对相应的槽位进行填充,即为“意图识别”任务和“槽位填充”任务。这两个任务旨在让计算机具备理解用户语言的能力,是具备语音交互能力的产品需要处理的重要任务。意图识别的目的是粗粒度地理解用户的目的,如针对表 1 中第一行的用户指令:“回放 11 月 3 号下午 4 点在 CCTV 11 播放的青春戏苑”,识别出意图是播放电视节目。槽位填充的目的是针对用户意图,细粒度地理解用户的具体要求,即根据抽取的槽位分析出用户想要播放的具体电视节目,相关的槽位包括“日期”“时间”“节目名称”等。

由于“意图识别”与“槽位填充”的准确率直接影响 AI 产品的性能与用户体验,如何有效且准确地识别用户意图以及抽取槽位一直是行业关注的热点。目前主流的一些基于预训练语言模型的联合建模方法^[1-3]都依赖于丰富的标注数据训练模型来完成“意图识别”和“槽位填充”。这些模型对数据集的标注样本要求很高,往往需要大量标注数据训练才能达到很好的效果,在数据稀缺的情况下模型效果则会显著下降。然而大量的标注数据意味着高昂的成本投入,特别是在真实业务场景下很难获取到高质量的标注数据。为了解决在真实语音交互业务中标注样本缺失的问题,本文提出一种新颖的基于小样本“意图识别”和“槽位填充”的串行解决方案。与联合建模方法不同,我们的方法采用串行架构,可以根据“意图识别”和“槽位填充”的特性,分别设计针对性更强的解决模块。

在“意图识别”任务中,当某一种意图的标注训练数据不足时,我们很难通过模型训练的方法识别该意图。由于数据缺失,训练出来的模型经常欠拟合,无法达到业务应用的标准。常见的补救方法是引入额外的标注数据来扩充训练集,然而这样的方法必定会引入额外的人工成本,如数据采集、数据标注、数据清洗等,因此并不是一种高效、简洁的方法。在本文中,针对“意图识别”任务中的数据稀疏问题,我们提出了

一种基于半监督学习的训练方法。我们的方法仅需要非常少的标注数据作为启动种子,就可以自动地识别小样本意图。在整个学习过程中,我们的方法不需要额外引入大量标注数据,且不需要人工干预,就可以高效、准确地完成“意图识别”任务。

在“槽位填充”任务中,当某一种意图下的槽位标注训练数据不足时,我们学习到的模型同样存在上述模型欠拟合问题,不能准确地对槽位进行填充。而且相较于“意图识别”任务中的数据标注,由于“槽位填充”任务中槽位种类繁多、形式各样,其数据标注过程更加复杂、耗时,获取成本极大。为了高效地对槽位进行填充,本文提出了一种基于迁移学习的训练方法。我们的方法在训练小样本槽位填充模型时,借鉴迁移了大样本槽位填充模型的能力,从而在小样本槽位数据不足的条件下,也可以训练出准确的槽位填充模型。同样的,在整个学习过程中,我们不需要引入额外标注数据,仅通过模型迁移学习,就可以得到准确的“槽位填充”模型。

综上,本文主要贡献有以下三点:

(1) 提出了一种针对语音交互的“意图识别”和“槽位填充”的串行解决方案,相比联合建模解决方案,我们的方案对小样本学习适用性更强。

(2) 针对“意图识别”任务中的数据缺失问题,提出了一种基于半监督学习的训练方法,不需要依赖额外的标注数据就可以提升模型准确性。

(3) 针对“槽位填充”任务中的数据缺失问题,提出了一种基于迁移学习的训练方法,凭借从大样本槽位数据中学习到的先验知识就可以提高小样本槽位填充模型的泛化能力。

本文提出的方法在 2021 年中国计算机学会大数据与计算智能大赛(CCF-BDCI)组委会与中国中文信息学会(CIPS)共同举办的全国信息检索挑战杯(CCIR Cup)中的“智能人机交互自然语言理解”赛道最终排名第一。

本文组织结构如下:第 1 节介绍相关工作;第 2 节详细介绍所提出的模型;第 3 节展示和分析实验结果;第 4 节总结全文。

1 相关工作

1.1 意图识别与槽位填充

早期的“意图识别”和“槽位填充”任务多是相互独立进行的,后来随着 BERT 等预训练模型的提

出, 用于联合建模“意图识别”和“槽位填充”任务的方法开始逐渐应用。按照实现方式的不同, 联合建

模方法又可分为隐式联合建模、显式联合建模以及预训练范式建模。

表 1 不同意图下样本槽位对照

指令	意图	槽位
回放 11 月 3 号下午 4 点在 CCTV11 播放的青春戏苑	播放电视节目	{“日期”: 11 月 3 号, “时间”: 下午四点, “频道”: CCTV11, “名称”: 青春戏苑}
把抽烟机的档位调大一些	家居控制	{“设备”: 抽烟机, “细节”: 调大一些, “命令”: 模式调节}
订三张明天早上飞成都的经济舱	旅行查询	{“日期”: 明天, “时间”: 早上, “目的地”: 成都, “查询类型”: 机票查询}
通知我在 5 月 30 日去打狂犬疫苗	闹钟提醒	{“日期”: 5 月 30 号, “记录”: 去打狂犬疫苗}
2019 年古装爱情电视剧小女花不弃的花絮播放一下	播放视频	{“名称”: “小女花不弃的花絮”, “日期”: 2019 年}

隐式联合建模: 隐式建模通过两个任务共享一个编码器模型实现, 建模过程没有明显的任务交叉, 经过编码的文本数据经过两个独立的解码器分别实现意图识别和槽位填充任务。Zhang 等人^[4]在 2016 年提出采用 RNN 网络作为编码器以学习意图和槽位间的依存关系。Liu 等人^[5]使用注意力机制和双向 RNN 替换了 RNN 作为编码器。Hakkani-Tür 等人^[6]又提出使用共享的 RNN-LSTM 框架用于联合方式建模。

显式联合建模: 显式建模是一种通过增加显式交互模块学习两个信息间的相互关系的建模方法。相比隐式联合建模, 这种方式具有显式可控、可解释等优点。Goo 等人^[7]在 2018 年提出了一种槽位门控联合模型, 这个模型让槽位信息可以根据学习到的意图表示的信息进行调节。Li 等人^[8]提出了一种具有意图增强的门控机制, 用于引导槽位填充的模型。

预训练范式建模: 预训练模型已经在各种 NLP 相关任务中取得了令人瞩目的成绩。Castellucci 等人^[1]使用了 BERT 模型用于联合意图分类和槽位识别任务, 在多个数据集上效果显著好于上述基于注意力的 RNN 和基于槽位门控等传统范式的模型。

与其他 NLP 任务相似, 由于预训练模型提供了较好的语义表征, 有助于提升模型在意图分类和槽位识别任务中的表现性能, 因此本文中提出的主要模型和方法都是基于预训练模型开展。

1.2 预训练模型

自 2018 年以来, 基于 Transformer^[9]的预训练语言模型 (PLM) 大规模兴起。BERT^[10]是使用最

广泛的预训练语言模型之一。受到 BERT 的启发, 已经出现了许多改进变体, 这里仅介绍本文所采用的三个预训练模型。RoBERTa^[11]在 PLM 任务中采用动态 Mask 替换静态 Mask, 并且使用更多的训练数据进行训练, 更为健壮。NEZHA^[12]在每一层计算隐状态的相互依赖时编码相对位置信息, 引入全词 Mask 技术, 训练过程中采用混合精度训练方式, LAMB 优化器极大地提高了训练速度。MacBERT^[13]在 PLM 任务中采用 Whole-Word Mask 和 N-gram Mask 策略, 被遮蔽词使用相似词替换, 减小了训练和微调阶段之间的差距, 在许多任务上取得了显著效果。

1.3 迁移学习

迁移学习旨在把来自相关领域(也称源域)的知识迁移至目标领域(也称目标域), 提高模型在目标领域的学习性能。根据文献[14], 迁移学习可以分为基于样本实例的迁移学习、基于特征的迁移学习、基于模型的迁移学习等策略。基于样本实例的迁移学习侧重于通过数据的调整来转换和实现知识的迁移, 针对源域中和目标域相似的样本赋予高权重, 不相似的样本赋予较低的权重。基于特征的方法通常采用特征转换策略, 将每一个原始特征转化为新的特征表示。根据文献[15], 以上两种方法又可以归为基于数据的迁移。基于模型的迁移主要目标是对目标域任务做出更准确的预测, 一般是通过预训练模型和参数共享实现。例如, Qin 等人^[16]提出动态聚合框架, 使用动态聚合函数细粒度建模领域之间的联系, 通过表示领域间相关性, 利用专家机制加权获得最终的领域私有特征, 具备领域共享特征的能力。

2 模型

本节介绍我们提出的基于小样本“意图识别”和“槽位填充”的串行解决方案模型,2.1 节介绍模型整体结构设计,2.2 节介绍基于半监督学习机制的小样本意图分类模型,2.3 节介绍基于迁移学习的小样本槽位填充模型。

2.1 模型结构

目前流行的语音交互方案主要对“意图识别”和“槽位填充”两个任务采取共享参数的联合建模方法,如 Goo 等人^[7]提出的槽位门控联合模型,Chen 等人^[3]提出的 Joint-BERT 模型,以及 Qin 等人^[2]提出的交互注意力模型。这类方法通过语义网络层参数的共享,虽然可以利用不同任务间的共享知识来增强模型,但却忽略了不同任务独有的特性,使模型很难兼顾两个任务优化方向而学

习到深度语义模型。尤其在小样本学习场景中,稀疏的小样本数据不足以支持语义网络学习好复杂的多任务共享知识。针对以上问题,本节提出了一种基于多任务串行的建模方案,使得我们可以利用不同任务的特性对不同任务独立学习建模,而后针对不同任务模块中的小样本问题,应用不同的训练方法独立学习。

本文提出模型的整体结构如图 1 所示。整体结构分为两大模块:意图分类(Intent Recognition)和槽位填充(Slot Filling),其中槽位填充模块进一步分为两个子模块,分别是大样本意图槽位填充和小样本意图槽位填充。我们第一步先将用户指令输入到意图分类模块获取用户指令的意图;第二步,根据第一步的结果判断用户的意图是否属于小样本意图,然后根据意图类别(大样本意图或小样本意图)选取对应的槽位填充模块来获取用户指令中的槽位。本文所采取的意图分类和槽位填充的基础构造如下:

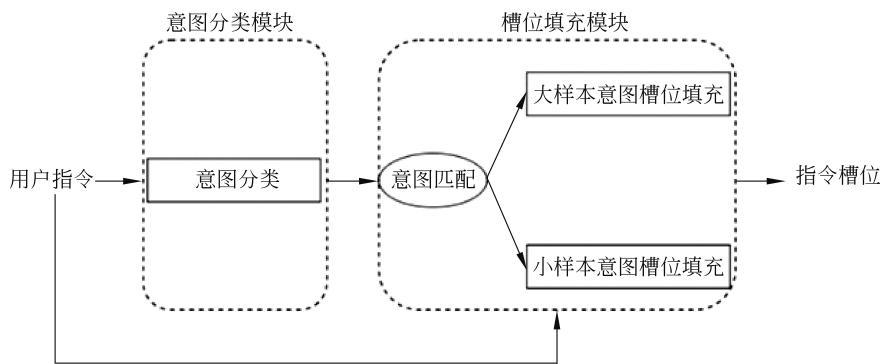


图 1 模型整体结构

意图分类模块: 该模块采用 BERT-base 结合前馈神经网络(FFN)进行意图分类。BERT-base 是由 12 层 Transformer 构成的编码器,向量维度是 768 维,FFN 由两个具有 ReLU 激活函数的多层感知机和一个线性层构成。给定用户指令 $X = \{x_1, x_2, \dots, x_l\}$,其中 l 代表句子长度。首先利用 BERT 提取 X 中每一个单词的上下文词向量,我们取 BERT 中最后一层的词向量,记为 $\mathbf{H} = \{h_1, h_2, \dots, h_l\} \in R^{l \times d}$, d 代表 BERT 的向量维度,也就是 768。我们用 \mathbf{H} 的平均池化结果作为整个句子的语义表示,记为 \mathbf{q} ,如式(1)、式(2)所示。

$$h_1, h_2, \dots, h_l = \text{BERT}\{x_1, x_2, \dots, x_l\} \quad (1)$$

$$\mathbf{q} = \frac{1}{l} \sum_{i=1}^l h_i \quad (2)$$

其中, $\mathbf{q} \in R^d$,最后将 \mathbf{q} 通过 FFN 计算分类的概率

p ,如式(3)所示。

$$p = \text{softmax}(\text{FFN}(\mathbf{q})) \quad (3)$$

其中, $p \in R^k$, k 表示类别的总数,取概率最大的作为预测的意图类别。

槽位填充模块: 槽位填充的两个子模块均采用 BERT-base 作为编码器,为了可以建模标签之间的依赖关系,提升模型在槽位识别上的效果,本文在 BERT 的输出词向量后引入 Bi-LSTM^[17] 结合 CRF^[18] 的网络模型。具体来说,我们将用户指令 X 输入到编码器, x_i 代表 X 中第 i 个单词的编码向量,最后 x_i 经 Bi-LSTM 转换处理得到最终表示向量为 \mathbf{h}_i 。

$$\vec{\mathbf{h}}_i = \text{LSTM}_f(\mathbf{w}_i, \vec{\mathbf{h}}_{i-1}; \theta_f) \quad (4)$$

$$\hat{\mathbf{h}}_i = \text{LSTM}_f(\mathbf{w}_i, \hat{\mathbf{h}}_{i-1}; \theta_b) \quad (5)$$

其中 $\mathbf{h}_i = \vec{\mathbf{h}}_i \oplus \hat{\mathbf{h}}_i$, \oplus 代表向量的拼接操作, θ_f, θ_b 表示前向 LSTM 和反向 LSTM 的参数, $\vec{\mathbf{h}}$ 和 $\hat{\mathbf{h}}$ 分别是前向和反向 LSTM 在第 i 个单词的隐藏状态; 然后, 我们将 \mathbf{h}_i 输入到 CRF 模型中计算所有可能的路径分数, 选择分数最大的作为预测的路径。

2.2 基于半监督学习机制的小样本意图分类

在图 1 中的意图分类模块中, 我们使用一个基于 BERT 的分类模型来识别用户指令的意图。然而在该模型的学习过程中, 由于日常用户指令的意图种类同时包含大样本的高频意图与小样本的低频意图, 使得我们在普通有监督的训练方式下学习到的模型仅能较好地识别出高频意图, 而对于低频意图往往识别得不尽如人意。因此, 本节提出一种基于半监督机制的学习策略, 使得模型可以在不引入额外人工标注样本的条件下大幅提升对小样本意图的识别能力。

针对含有小样本的意图分类问题, 我们的解决思路是依靠半监督学习的方法, 使得模型可以自动地从未标注的数据集中获取高置信度的小样本数据, 而后将其打上伪标签重新代入模型进行学习, 重复整个过程直至模型收敛, 具体的方法流程如下:

第一步, 利用已有大样本标注数据与少量小样本标注数据学习一个意图分类模型, 并利用该模型对未标记的用户指令进行意图分类和置信度打分, 其中置信度通过人工干预, 选取高置信度数值, 有助于在迭代过程中获得好的效果。在初始轮次, 该模型仅使用有限少量小样本数据训练, 而在余下轮次, 该模型会额外引入伪标签小样本数据进入训练集。

第二步, 对于从第一步中获取的样本意图以及置信度, 如果一个用户指令被预测为小样本意图的置信度大于 δ_1 并且被预测为任何大样本意图的置信度小于 δ_2 , 我们就认为该指令的意图是小样本意图, 并将其打上小样本的伪标签置入训练数据中。在初始轮次, 由于模型小样本分类能力还比较差, 可能会错识别一些小样本数据, 但随着模型迭代, 该类错误样本会被正确识别并被移出训练集。

第三步, 重复第一、二步, 直至模型收敛。我们每一轮都会对未标注的数据意图进行重新预测, 产生新的伪训练数据, 当模型预测结果与上一轮预测结果一致时, 视为模型已经收敛。值得注意的是, 在

上述流程中, 我们主要目的是通过加入伪标签的小样本数据解决高频意图与低频意图的数据不平衡问题, 所以我们并不对大样本训练数据进行额外处理。算法将两类小样本统一概括为小样本问题表述, 具体伪代码流程图如下:

Algorithm 1 基于半监督机制的意图分类模型学习方法

Input: 有标注大样本集 D , 有标注小样本集 R , 无标注样本集 U , 筛选样本的置信度阈值 δ_1, δ_2 .

Output: 多分类模型 M

```

1:  $R' = \emptyset$ 
2: while  $R \neq R'$  do
3:    $R' = R$ 
4:    $M = \text{LEARN}(D \cup R)$  // 利用  $D$  和  $R$  训练意图分类模型  $M$ 
5:   for each  $u$  in  $U$  do
6:      $P = M(u)$  // 预测样本  $u$  的概率分布
7:     //  $c$  代表小样本类别标签的索引
8:     if  $P[c] > \delta_1$  AND  $\max(\bigcup_{i \neq c} \{P[i]\}) < \delta_2$  then
9:        $R = R \cup \{u\}$  // 将满足条件的小样本加入到训练集
10:      end if
11:    end for
12:  end while
13: return  $M$ 

```

2.3 基于迁移学习的小样本槽位填充

如图 1 所示, 我们的模型中槽位填充模块包含两个子模块, 分别用于处理大样本意图的槽位填充和小样本意图的槽位填充, 两个子模块均采用 BERT+BiLSTM+CRF 的网络结构。当填充大样本意图的槽位时, 由于大样本的槽位标注数据充足, 我们可以直接通过上述网络模型学习到较为精准的槽位填充模型用于槽位填充。然而当填充小样本意图的槽位时, 由于小样本的槽位标注数据不足, 采取直接学习的方法不能有效地获取满意的槽位填充模型, 因此我们需要额外应用针对小样本的学习方法来训练模型。相较于“意图识别”任务, 由于“槽位填充”任务中槽位种类繁多、组合各样、结构复杂, 所以类似 2.2 节所提出的针对数据分类任务的半监督学习策略很难适用于该任务, 因此本文针对“槽位填充”任务的特性, 提出了一种基于迁移学习的训练方法, 在不引入任何额外有标注数据或无标注数据的情况下, 仅凭少量小样本数据也可以训练出准确的槽位填充模型。

该模块主要思路如图 2 所示, 对于小样本槽位填充任务, 我们可以利用迁移学习的思想, 利用小样本标注数据对已训练好的大样本模型, 进行适应性调整便可获得强健的小样本槽位填充模型。我们具体的学习过程概括如下, 假设对于 N 个大样本意图, 我们首先学习 N 个大样本意图的槽位

填充模型；然后，对于这 N 个学习好的大样本模型，我们截取其输出层之前的网络结构及权重用于模型迁移；接下来，对于某个小样本槽位填充任务，我们分别在这 N 个迁移的网络模型上，利用有限的小样本对模型进行微调训练，得到 N 个新的

针对小样本意图的槽位填充模型。最后，在槽位预测时，我们利用这 N 个新的小样本槽位填充模型，对某一用户指令，分别输出 N 个槽位预测结果，综合这 N 个结果投票，取票数最多的结果作为最终的结果。

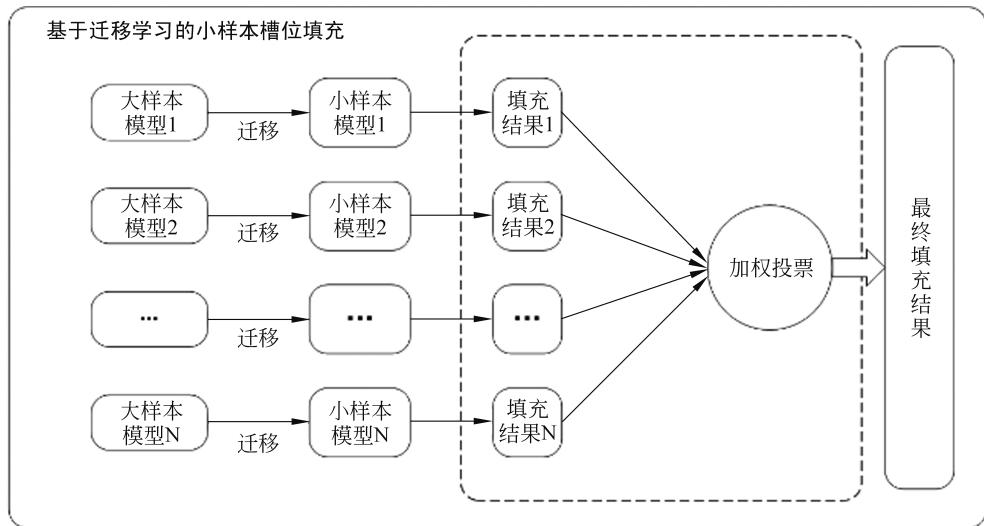


图 2 基于迁移学习的小样本槽位填充模块示意图

值得注意的是，在上述投票过程中，由于 N 个模型的先验知识并不相同，各模型预测出的结果的置信度亦不相同，所以将 N 个模型的预测结果进行等权投票的方法并不合理。显然，如果迁移之前的来源大样本数据与之后的目标小样本数据更相似，则学到的模型置信度更高；如果来源数据与目标数据间的差异偏大，则学到的模型置信度偏低。如表 1 所示，如在学习小样本意图“播放电视节目”下的槽位时，如果来源数据是大样本意图“播放视频”，则此时模型的预测结果应该给予较高的置信度，因为二者数据较相似；相反，如果来源数据属于“家居控制”意图，与“播放电视节目”意图的数据差异过大，则应给予较低的置信度。因此，本文在最终模型投票阶段，并不采取等权投票，而是按槽位模型的来源数据与目标数据的相似度的大小分配投票权重大小。

本文算法最终考虑了两种槽位相似度，作为衡量模型投票权重大小的标准。第一种是槽位本身的语义相似度，主要是从槽位语义的角度出发，衡量两个槽位的语义相似度。如“日期”和“时间”的语义相似度高，而与“设备”的语义相似度低。具体地，我们利用 BERT 预训练的语义向量表达两个槽位名的语义向量，而后直接计算两向量之间的 Cosine 相似度作为二者的语义相似度，如式(6)所示。

$$\text{Sim}_{\text{name}}(z, z') = \text{Cosine}(\mathbf{W}_z, \mathbf{W}_{z'}) \quad (6)$$

其中， z 和 z' 代表两个不同的槽位， \mathbf{W}_z 和 $\mathbf{W}_{z'}$ 分别是两个槽位对应的语义向量，Cosine 代表两个向量之间的余弦相似度。此外本文考虑的第二种相似度为槽位所包含内容的相似度，主要是根据两槽位是否包含相同的槽位值，来判断两个槽位是否相似。如果两个槽位有越多相同的槽位值，则说明它们越相似，反之越不相似。如“天气查询”意图下的槽位“城市”和“旅行查询”意图下的槽位“目的地”、“出发地”下都有大量相同的城市名如北京、桂林等，所以“城市”和“目的地”、“出发地”应为相似槽位；相反的，与其它不包含任何地点的意图，例如播放视频没有相同的槽位值，所以不相似。具体的，我们用 Jaccard 相似度表示上述相似度，如式(7)所示。

$$\text{Sim}_{\text{value}}(z, z') = \frac{|V_z \cap V_{z'}|}{|V_z \cup V_{z'}|} \quad (7)$$

其中， V_z 和 $V_{z'}$ 分别代表槽位 z 和 z' 对应的槽位值的集合。综合以上两种相似度，我们使用二者的加权结果作为最终的槽位相似度，表示为 $\text{Sim}_{\text{slot}}(z, z')$ ，计算如式(8)所示。

$$\text{Sim}_{\text{slot}}(z, z') = \alpha * \text{Sim}_{\text{name}}(z, z') + \beta * \text{Sim}_{\text{value}}(z, z') \quad (8)$$

其中， α 和 β 为权重系数。本节所述完整算法流程如下：

Algorithm 2 基于迁移学习的槽位填充方法

Input: 目标小样本槽位 z , 目标小样本标注数据集 D , 用户指令 S , 已从大样本意图数据上学习好的槽位填充模型集合 H 。

Output: 槽位标记序列 R

```

1:  $H' = \{\}$ 
2:  $R = [0, 0, \dots, 0]$  //R 的长度即为 S 的长度
3: //学习过程
4: for each  $h_{z'}$  in  $H$  do
5:    $h_{z'} = \text{LEARN}(h_{z'}, D)$ //基于已学好的大样本槽位  $z'$  的模型, 进一步在目标小样本数据集  $D$  上学习小样本槽位语义特征
6:    $H' = H' \cup \{h_{z'}\}$ 
7: end for
8: //预测过程
9: for  $i = 1$  to  $\text{len}(S)$  do
10:    $Y_z = [0, 0, \dots, 0]$ // $Y_z$  为  $S[i]$  关于槽位  $z$  的标签预测的概率分布
11:   for each  $h_{z'}$  in  $H'$  do
12:      $Y'_{z'} = \text{PREDICT}(h_{z'}, S[i])$ 
13:      $Y_z = Y_z + Y'_{z'} * \text{Sim}_{\text{slot}}(z, z')$ //根据槽位相似性预测当前的槽位
14:   end for
15:    $R[i] = \text{argmax}_j Y_z[j]$ // $R[i]$  为  $S[i]$  被预测的概率最大的标签
16: end for
17: return  $R$ 

```

3 实验

3.1 数据集介绍

本文采用 2021 年中国计算机学会大数据与计算智能大赛(CCF-BDCI)组委会与中国中文信息学会(CIPS)共同举办的全国信息检索挑战杯(CCIR Cup)中的“智能人机交互自然语言理解”^①。数据集包含用户与音箱等智能设备进行单轮对话的文本数据, 包括共 11 个意图类别, 47 个槽位类型。训练集有 9 100 条, 每一个样本如表 2 所示, 包含意图和槽位。该赛题的评测任务共有三个子任务(表 2)。

3.2 实验参数

意图分类模块的预训练模型采用 MacBERT、

表 2 赛题评测任务

子任务	数据分布
基础任务	包含 9 个类别, 每一个类别有对应的槽位以及 1 000 条样本
小样本任务	包含 2 个类别, 每一个类别有对应的槽位以及 50 条样本
域外检测任务	无训练数据提供, 将除了以上 11 个类别的数据视为域外样本。域外类别没有槽位

NEZHA、RoBERTa 投票集成, 槽位填充模块的预训练模型采用 MacBERT。句子最大长度设置为 256, 优化算法为 AdamW, 初始学习率设置为 2e-5, 半监督学习轮次设置为 5、两个小样本意图预测置信度参数 δ_1 和 δ_2 分别设置为 0.95 和 0.15, 槽位填充算法中两种相似度被视为同等重要影响因子, 权重系数 α 和 β 均设置为 0.5, 对抗训练的扰动参数设置为 0.05, 批次 batch 设置为 64, 训练 epoch 设置为

20。以上均为目前针对本实验数据所得的最优参数。

3.3 实验结果

首先我们从训练集中按照 10 : 1 比例划分训练集和测试集, 对比离线模式下不同预训练模型在槽位填充任务上的表现, 实验结果如表 3 所示。

表 3 预训练模型对比实验

模型	意图分类	槽位识别
BERT + Naive	0.933	0.874
RoBERTa + Naive	0.935	0.871
ALBERT + Naive	0.933	0.876
NEZHA + Naive	0.940	0.887
MacBERT + Naive	0.951	0.903

注: Naive 指代用 BERT 最后一层表示做意图分类, 用 Transformer 做槽位的序列标注。

根据实验结果所示, 使用 MacBERT 预训练模型做下游意图分类和槽位识别任务的效果好于其他模型。相比其他 BERT 模型, 除了使用全词掩码策略

外, MacBERT 在预训练阶段采用相似词替换的掩码策略, 以降低预训练和微调阶段两者之间的差距, 使

^① <https://www.datafountain.cn/competitions/511>

得 MacBERT 更加贴合下游任务,因此本文在两个任务中均以 MacBERT 作为我们的预训练模型。

在上述确定预训练模型中,我们仅使用大样本数据集进行实验以确定后续任务的基础预训练

模型。为适配小样本实验结果,我们针对比赛任务对全部训练数据进行学习,并提交模型预测结果到线上验证。表 4 中提供了实验模型及线上提交结果。

表 4 模型评测分数对照

序号	模型	准确率
(1)	MacBERT + JointLoss	0.562
(2)	MacBERT + Naive	0.607
(3)	MacBERT + BiLSTM	0.616
(4)	MacBERT + BiLSTM + CRF	0.622
(5)	MacBERT + BiLSTM + CRF + SFSL	0.754
(6)	MacBERT + BiLSTM + CRF + SFSL + TFSL	0.848
(7)	MacBERT + BiLSTM + CRF + SFSL + TFSL + FGM	0.853

通过对比模型(1)和模型(2),我们发现基于意图分类和槽位识别的串行模型结构在模型效果上要好于并行模型架构。通过 JointLoss 联合学习意图分类和槽位识别任务无法协调其中一类任务的优化,因此在我们的模型中确定使用串行模型架构的学习方式。

在模型结构上,我们分别对(2)、(3)、(4)三种模型结构进行了对比实验。通过在 MacBERT 基础上添加 BiLSTM 增强语义特征提取能力,我们获得了约 1 个百分点的提升。除此之外,加入条件随机场后,通过对标签预测结果加入条件约束,模型对预测结果得到更加准确的结果。

在提升小样本适应能力、模型迁移能力和鲁棒性方面,我们分别向模型中注入基于半监督学习机制的小样本意图分类算法、基于迁移学习的小样本槽位填充算法和对抗训练方法。在基础模型(4)的基础上,将基于半监督学习机制的小样本意图分类

算法(SFSL)引入,即模型(5),推理结果获得“质”的提升,其原因在于该算法对小样本意图分类准确率提升显著,并且由于我们采取了串行模型架构,意图识别准确率的提升也减少了之前因意图误分导致的槽位填充错误。当结合基于半监督学习机制的小样本意图分类算法后,再引入基于迁移学习的小样本槽位填充算法(TFSL),即在模型(5)的基础上对小样本槽位识别进行改进得到模型(6),由于我们将大样本槽位模型的能力通过迁移学习泛化到小样本槽位模型上,使得小样本槽位填充模型的准确率得到显著提升,进一步显著地提升了模型整体的准确性。最后,模型(7)通过加入对抗训练(FGM)^[19] 和多模型加权投票融合,取得最优的线上评测结果。

3.4 实验分析

为了更好地说明我们方法的有效性,本文给出了具体的实例做结果对比分析,见表 5 和表 6。

表 5 与联合建模策略在意图分类效果上的对照实例

文本	联合建模	独立建模
厦门飞天津的飞机已经起飞了吗	旅行查询	其他
看一下今天的南海海洋天气预报,调到南海	天气查询	电视节目播放
我要循环播放张籍的秋思	音乐播放	音频播放

表 6 大样本槽位迁移建模在小样本槽位填充效果上的对照实例

文本	相关大样本槽位迁移建模	不相关大样本槽位迁移建模
播放童话故事 谁是最幸运的 英文版的	{“来源”:“收音机收听”,“目标”:“音频播放”,“名称”:“谁是最幸运的”,“语言”:“英语”,“标签”:“童话”}	{“来源”:“日历查询”,“目标”:“音频播放”,“名称”:“是最幸运的”,“语言”:“英语”,“标签”:“Null”}

续表

文本	相关大样本槽位迁移建模	不相关大样本槽位迁移建模
播放香港喜剧 我爱钟无艳	{“来源”:“视频播放”,“目标”:“电影电视剧播放”,“名称”:“我爱钟无艳”,“标签”:“喜剧”,“区域”:“香港”}	{“来源”:“家具控制”,“目标”:“电影电视剧播放”,“名称”:“钟无艳”,“标签”:“喜剧”,“区域”:“Null”}
帮我搜索一下 CCTV 2 的一 槌定音	{“来源”:“电影电视剧播放”,“目标”:“电视节目播放”,“频道”:“CCTV 2”,“名称”:“一槌定音”}	{“来源”:“天气查询”,“目标”:“电视节目播放”,“频道”:“CCTV”,“名称”:“一槌定音”}

从表 5 可以看出,对于联合建模的方式,由于同时建模意图与槽位信息,因此当句子中出现某些相似的槽值时,可能会混淆意图的语义。如“我要循环播放张籍的秋思”这个句子中含有“作者、名称”槽位,因此很容易和“音频播放”意图混淆。如表 5 所示,尤其在小样本意图上,如“其他”“电视节目播放”“音频播放”,联合建模的方法很容易混淆不同意图的用户指令。而本文提出的模型,通过独立建模的方式与半监督的学习方法,可以正确识别小样本意图。

在表 6 中,我们通过从不同大样本模型上迁移小样本模型,来说明利用槽位相似度做预测加权投票的必要性。以“播放香港喜剧我爱钟无艳”为例,该文本真实意图是“电影电视剧播放”,真实槽位为“{“名称”:“我爱钟无艳”,“标签”:“喜剧”,“区域”:“香港”}”。如果我们迁移的大样本槽位模型来源是在“视频播放”意图下学习到的,则由于“视频播放”数据与“电影电视剧播放”数据的样本相似性高,所以迁移学习后的小样本模型可以正确识别槽位;反之,如果我们迁移的大样本槽位模型来源是在“家居控制”意图下学习到的,则由于“家居控制”数据与“电影电视剧播放”数据的样本相似性低,所以迁移学习后的小样本模型很难正确识别槽位。因此我们所提出的方法,在迁移各大样本模型后,通过式(7)加权投票,减弱数据相似度低的迁移模型权重与加强数据相似度高的迁移模型权重,以获得更为可信的预测结果。

4 结束语

本文提出的模型主要用于解决在真实语音交互业务中标注样本缺失的问题,提出一种新颖的基于小样本“意图识别”和“槽位填充”的串行解决方案。与联合建模方法不同的是,我们的方法采用串行架构,可以根据“意图识别”和“槽位填充”的特性,分别设计针对性更强的解决模块。基于本文所提出的模

型,我们在 2021 年中国计算机学会大数据与计算智能大赛(CCF-BDCI)组委会与中国中文信息学会(CIPS)共同举办的全国信息检索挑战杯(CCIR Cup)中的“智能人机交互自然语言理解”赛道取得了第一名的成绩。

参考文献

- [1] CASTELLUCCI G, BELLOMARIA V, FAVALLI A, et al. Multi-lingual intent detection and slot filling in a joint BERT-based model[J]. arXiv preprint arXiv: 1907.02884, 2019.
- [2] QIN L, LIU T, CHE W, et al. A cointeractive transformer for joint slot filling and intent detection[C]// Proceedings of International Conference on Acoustics, Speech and Signal Processing, 2021: 8193-8197.
- [3] CHEN Q, ZHUO Z, WANG W. BERT for joint intent classification and slot filling[J]. arXiv preprint arXiv: 1902.10909, 2019.
- [4] ZHANG X, WANG H. A joint model of intent determination and slot filling for spoken language understanding[C]// Proceedings of International Joint Conference on Artificial Intelligence, 2016: 2993-2999.
- [5] LIU B, LANE I. Attention-based recurrent neural network models for joint intent detection and slot filling [J]. arXiv preprint arXiv: 1609.01454, 2016.
- [6] HAKKANI-TÜR D, TÜR G, CELIKYILMAZ A, et al. Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM [C]// Proceedings of International Speech Communication Association, 2016: 715-719.
- [7] GOO C W, GAO G, HSU Y K, et al. Slot-gated modeling for joint slot filling and intent prediction [C]// Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018: 753-757.
- [8] LI C, LI L, QI J. A selfattentive model with gate mechanism for spoken language understanding[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018: 3824-3833.

- [9] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of Advances in Neural Information Processing Systems, 2017: 5998-6008.
- [10] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pretraining of deep bidirectional transformers for language understanding[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171: 4186.
- [11] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach[J]. arXiv preprint arXiv: 1907.11692, 2019.
- [12] WEI J, REN X, LI X, et al. NEZHA: neural contextualized representation for Chinese language understanding[J]. arXiv preprint arXiv: 1909. 00204, 2019.
- [13] CUI Y, CHE W, LIU T, et al. Revisiting pre-trained models for Chinese natural language processing[J]. arXiv preprint arXiv: 2004.13922, 2020.
- [14] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(10): 1345-1359.
- [15] ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning[J]. Proceedings of the IEEE, 2020, 109(1): 43-76.
- [16] QIN L, XU X, CHE W, et al. Dynamic fusion network for multi-domain end-to-end task-oriented dialog[J]. arXiv preprint arXiv: 2004.11019, 2020.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [18] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]//Proceedings of the 18th International Conference of Machine Learning, 2001: 282-289.
- [19] GOODFELLOW I, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv: 1412.6572, 2014.



孙相会(1997—),硕士研究生,主要研究领域为自然语言处理。

E-mail: sun64410822@163.com



窦辰晓(1988—),通信作者,博士,主要研究领域为数据挖掘、自然语言处理。

E-mail: chx.dou@gmail.com



苗德强(1997—),硕士研究生,主要研究领域为自然语言处理。

E-mail: tbxunrhx95@163.com