

## 基于深度学习的口语理解联合建模算法综述<sup>\*</sup>

魏鹏飞, 曾碧, 汪明慧, 曾安

(广东工业大学 计算机学院, 广东 广州 510006)

通信作者: 曾碧, E-mail: zb9215@gdut.edu.cn



**摘要:** 口语理解是自然语言处理领域的研究热点之一, 应用在个人助理、智能客服、人机对话、医疗等多个领域. 口语理解技术指的是将机器接收到的用户输入的自然语言转换为语义表示, 主要包含意图识别、槽位填充这两个子任务. 现阶段, 使用深度学习对口语理解中意图识别和槽位填充任务的联合建模方法已成为主流, 并且获得了很好的效果. 因此, 对基于深度学习的口语理解联合建模算法进行总结分析具有十分重要的意义. 首先介绍了深度学习技术应用到口语理解的相关工作, 然后从意图识别和槽位填充的关联关系对现有的研究工作进行分析, 并对不同模型的实验结果进行了对比分析和总结, 最后给出了未来的研究方向及展望.

**关键词:** 意图识别; 槽位填充; 注意力机制; 胶囊网络; 任务对话系统; 深度学习

**中图法分类号:** TP391

中文引用格式: 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>

英文引用格式: 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). <http://www.jos.org.cn/1000-9825/6385.htm>

## Survey on Joint Modeling Algorithms for Spoken Language Understanding Based on Deep Learning

WEI Peng-Fei, ZENG Bi, WANG Ming-Hui, ZENG An

(School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** Spoken language understanding is one of the hot research topics in the field of natural language processing. It is applied in many fields such as personal assistants, intelligent customer service, human-computer dialogue, and medical treatment. Spoken language understanding technology refers to the conversion of natural language input by the user into semantics representation, which mainly includes 2 sub-tasks of intent recognition and slot filling. At this stage, the deep modeling of joint recognition methods for intent recognition and slot filling tasks in spoken language understanding has become mainstream and has achieved sound results. Summarizing and analyzing the joint modeling algorithm of deep learning for spoken language learning is of great significance. First, it introduces the related work to the application of deep learning technology to spoken language understanding, and then the existing research work is analyzed from the relationship between intention recognition and slot filling. The experimental results of different models are compared and summarized. Finally, the challenges that future research may face are prospected.

**Key words:** intention recognition; slot filling; attention mechanism; capsule network; task dialogue system; deep learning

近年来, 随着计算机和互联网的蓬勃发展, 大规模并行计算成为可能, 同时也积累了海量的数据. 作为人工智能(artificial intelligence, AI)的核心技术, 深度学习开始借助强大的计算能力和海量数据, 构建规模更大的神经网络, 在计算机视觉、语音识别、自然语言处理、自动驾驶等领域取得了领先的成绩<sup>[1]</sup>. 对话系统的研

\* 基金项目: 国家自然科学基金(61772143); 广东省自然科学基金(2018A030313868); 广东省产学研重大专项(2016B010108004)

收稿时间: 2020-08-04; 修改时间: 2020-09-30, 2021-04-16; 采用时间: 2021-05-25; jos 在线出版时间: 2021-08-03

究也开始由早期的基于规则和基于统计方法的建模方式, 转变为以运用深度学习技术为主, 采用神经网络模型去建模各个模块<sup>[2]</sup>.

对话系统一般包含 5 大模块(López-Cózar 等人<sup>[3]</sup>、侯丽仙等人<sup>[4]</sup>将对话系统分成 5 个模块: Automatic Speech Recognition (ASR), Spoken Language Understanding (SLU), Dialogue Management (DM), Natural Language Generation (NLG), Text-to-Speech synthesis (TTS)), 本文采用 Chen 等人<sup>[5]</sup>提出的划分方式, 将其分为 4 个部分: 口语理解(spoken language understanding, SLU)(又称为自然语言理解(natural language understanding, NLU)<sup>[6]</sup>)、对话状态跟踪(dialogue state tracking, DST)、对话策略学习(dialogue policy learning, DPL)、自然语言生成(natural language generation, NLG), 如图 1 所示.

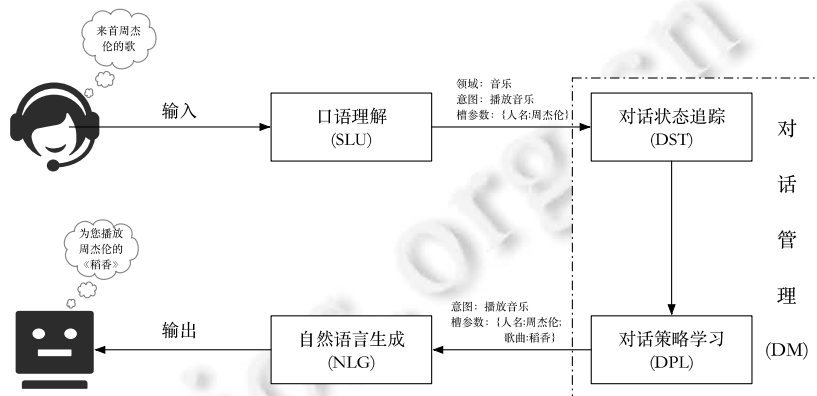


图 1 对话系统架构

SLU 是对话系统中关键的功能模块, 它的好坏直接影响整个对话系统的性能. 给定用户话语, 口语理解将话语映射为结构化的语义表示. 如图 2 所示, 用于语义表示的流行模式是对话行为, 由意图和槽位组成. 意图是话语的类别, 例如“查找和通知”, 表明话语的功能. 槽位是话语中提到的特定于任务的实体, 例如“周杰伦和稻香”, 给出完成任务所需的实体约束.

基于这种对话行为的结构, SLU 的任务可以进一步分解为两个任务: 意图识别(intent recognition, IR)和槽位填充(slot filling, SF). 前者 IR 通常被表示为以话语为输入的意图分类任务, 而 SF 通常被视为序列标注问题. 图 2 中, “查找周星驰最近的喜剧”是电影相关话语示例, 其中, “find\_movie”是意图分类的结果, “O O B-dir I-dir I-dir B-date I-date O B-genre I-genre”是序列标注的结果. 而 IOB 是一种标准的标注格式, 有利于语义元素的提取: 导演名称(director name)、日期(date)和类型(genre).

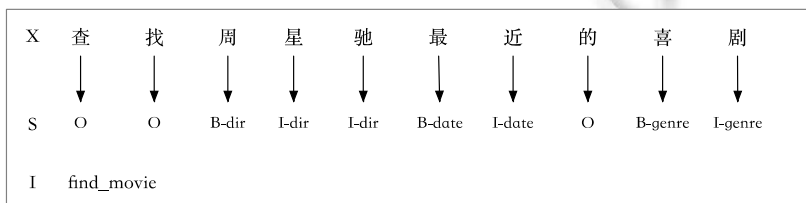


图 2 示例话语带有 IOB 格式(S)和意图(I), B-dir 和 I-dir 表示导演名称的语义槽注释

SLU 可以看作是对语言的理解. 语言主要有两种表现形式: 书面语和口语. 书面语是用于书写的, 口语是用于口头表达的. 书面语是在口语的基础上演变和发展而来的, 书面语规范、理性、结构严密、表述清晰; 而口语较为直接、感性、往往不假思索, 脱口而出, 用词范围比较狭窄, 可以重复、脱节、颠倒、补说, 也可以有大量的语气词和“这个”“那个”之类的带个性特征的习惯性的废话. 口语首先要经过自动语音识别技术(automatic speech recognition, ASR)转成口语文本, 此文本常常伴随着一些识别错误. 因此, SLU 模块在处理口

语文本时更加困难.

早期研究中,意图识别和槽位填充是分开进行建模的.在运行时,这种任务的流水线处理方式将导致错误从一个任务转移到下一个任务.此外,各个领域模型之间没有数据或功能共享,无法充分地利用所有的监督信号.于是就出现了联合建模方法.意图识别和槽位填充的联合建模原因可总结为以下 3 点.

- (1) 在 SLU 系统中,这两个任务通常同时出现.例如,要完成用户对于“预定后天从广州飞往北京的航班”的需求,则需要解析出用户意图:预定航班;和槽位:时间、出发地和目的地.
- (2) 一项任务的信息可以在另一项任务中使用,相互促进<sup>[7]</sup>.例如,如果话语的意图是查找航班,则它很可能包含出发地和目的地;反之亦然.
- (3) 意图和槽位标签作为用户行为的语义表示,应该相互共享知识.例如,预定航班和查找航班都需要同时提供出发地和目的地.

近年来,基于深度学习构建的意图识别和槽位填充的联合模型取得了很大的进展.大致可以分为 3 类:意图识别和槽位填充无关联的联合建模、意图识别关联槽位填充的联合建模以及意图识别和槽位填充双向关联的联合建模.无关联的联合建模方法并没有显式地构建两个任务之间的关联关系,Guo 等人<sup>[8]</sup>提出使用递归神经网络和给定的语法结构来学习句子的语义信息,根节点的输出向量和融合了上下文单词信息的三路径向量,分别用来预测意图和槽位标签.Liu 等人<sup>[9]</sup>使用双向 LSTM 学习句子特征,简化了树状结构,消除了生成语法树所带来的时间和人工成本.最终的输出向量一方面直接用作意图标签的预测,另一方面直接输入到带有注意力机制的单向 LSTM 网络中去预测槽位标签.它们都只是在编码器层进行了共享表示学习,虽然都是联合建模,但是两个任务的输出并未进行关联.之后,随着注意力机制应用在编解码模型上取得了很好的效果<sup>[10,11]</sup>,研究者开始探索将意图的信息关联到槽位填充任务中,这些文献<sup>[12-14]</sup>使用注意力机制学习句子上下文信息,可以很好地实现并行计算,比 LSTM 更加高效.门控机制(gate mechanism)/掩码机制(mask mechanism)利用意图信息来增强槽位标签的预测,这种单向的任务之间的信息传递,促使两个任务的性能都得到了显著的提升.反之,槽位信息对意图识别任务也有一定帮助,为了建立这两种任务之间的关联关系,E 等人<sup>[15]</sup>设计了 SF-ID 网络:SF 子网和 ID 子网,SF 子网引入相关因子将意图信息应用于槽位填充任务,而 ID 子网也计算一个相关因子将槽位信息用于意图识别任务,然后,这两个子网通过迭代机制进行训练,以帮助彼此相互促进.目前,对于以上的研究成果还没有很好的总结性工作.虽然侯丽仙等人<sup>[4]</sup>从口语理解的独立建模和联合建模方面对之前的相关研究做了总结,但是并没有考虑到模型内部深层次的关联关系和多域数据集评测指标的对比.而且,近年来又出现许多架构更加新颖、性能更优的一些模型.本文围绕口语理解的深度学习建模方法,重点从意图识别和槽位填充的关联关系上梳理了相关研究成果,给出了相关数据集及其评测指标,并从实验指标上对不同模型进行分析和总结,以帮助指导未来的研究和发展.

本文第 1 节介绍相关工作与有关设计问题.第 2 节介绍口语理解涉及的深度学习技术.第 3 节给出口语理解任务普遍采用的两个数据集.第 4 节从意图识别和槽位填充的关联关系上梳理了相关研究成果.第 5 节从实验指标上对不同模型进行对比分析.第 6 节总结了论文下载地址与源码链接.第 7 节对未来研究方向进行展望.

## 1 相关工作

关于口语理解的研究已有很长的历史,这种技术是在 90 年代从某些呼叫分类系统<sup>[16]</sup>(不限成员名额的口语响应来自动路由电话的任务)和 ATIS 项目<sup>[17]</sup>中出现的.可将口语理解任务定义为  $p_{intent}(d|x_1, x_2, \dots, x_n)$  和  $p_{slot}(y_1, y_2, \dots, y_n|x_1, x_2, \dots, x_n)$ , 其中,  $d$  表示意图类别,  $y_1$  到  $y_n$  是输入话语  $[x_1, x_2, \dots, x_n]$  中对应每个标记的槽位标签,  $n$  是输入话语标记的数量.本节中,我们主要介绍有关意图识别和槽位填充的一些典型工作,它们是口语理解的核心任务<sup>[5,18]</sup>.

意图识别可被看作是分类任务.早期的传统方法是将  $n$ -gram 作为通用实体特征,如位置和日期<sup>[7]</sup>,这种类型的方法局限于输入空间的维数.另一种流行的方法是在标记的训练数据上训练机器学习模型<sup>[19,20]</sup>,例如

SVM<sup>[21]</sup>、Adaboost<sup>[22]</sup>和改进的 KNN 方法 MLKNN<sup>[23]</sup>来改善意图识别. 基于神经网络架构的方法在意图识别任务中也表现出良好的性能. 深度信念网络(DBN)最早用于呼叫路由分类<sup>[24]</sup>, RNN<sup>[25]</sup>, IndRNN- Attention<sup>[26]</sup>和 Hybrid ELMo<sup>[27]</sup>在意图分类任务上表现优异.

对于槽位填充, 通常看作序列标注任务. 传统的方法是基于条件随机场(CRF)体系架构, 它具有很强的序列标记能力<sup>[28]</sup>, 并结合最大熵模型用于信息提取与分割<sup>[29]</sup>. 近年来, 基于神经网络及其扩展的方法在槽位填充任务中表现出出色的性能, 并且优于传统的 CRF 模型. 例如, Yao 等人<sup>[30]</sup>提出在标准的递归神经网络语言模型中将单词作为输入, 然后预测槽位标签. Yao 等人<sup>[31]</sup>通过使用过度特征和 CRF 的序列级优化标准对输出标签的依赖关系进行显式建模, 从而改进了传统的 RNN. Mesnil 等人<sup>[32]</sup>尝试使用双向和混合 RNN 进行槽位填充. Yao 等人的另一篇文章<sup>[33]</sup>为此任务引入了 LSTM 结构, 并比 RNN 获得了少量改进. 此外, 基于注意力的模型在 NLP 领域获得成功之后, Simonnet 等人<sup>[34]</sup>将基于注意力的编码器/解码器(encoder-decoder)应用于槽位填充任务, 而没有使用 LSTM 单元. Zhu 等人<sup>[35]</sup>提出了一种具有聚焦机制的 BLSTM-LSTM 模型来处理序列问题.

研究发现: 以上独立建模方法常常会引起错误传播问题; 而且这两个任务都是同时出现的, 相互之间还有一定的关联关系. 为了解决以上问题, 联合建模方法便被提出, 并已成为当前研究的主要方向. Hakkani-Tur 等人<sup>[36]</sup>首先提出了使用 RNN 的联合工作, 以学习意图和槽位之间的相关性. Zhang 和 Wang<sup>[7]</sup>使用双向 GRU 网络用于联合建模槽位填充和意图识别. Liu 等人<sup>[9]</sup>提出了一种基于注意力的神经网络, 以对两个任务进行联合建模. 以上这些模型都属于管道模型. 但是这些联合模型没有明确地为槽位和意图信息建模, 而是通过共享参数来考虑它们在两个任务之间的相关性. 近年来, 一些联合模型探索并使用了用于槽位填充的意图信息. Goo 等人<sup>[12]</sup>利用门控机制作为特殊的门功能来模拟意图识别和槽位填充之间的关系. Li 等人<sup>[13]</sup>提出了意图增强门控机制, 以利用槽位和意图之间的语义相关性. Qin 等人<sup>[37]</sup>使用堆栈传播框架将单词级的意图识别信息直接输入到槽位填充中. 同时也出现了将意图信息和槽位信息双向关联的模型, Wang 等人<sup>[38]</sup>提出了 Bi-Model 以考虑意图和槽位之间交叉影响. E 等人<sup>[15]</sup>引入了 SF-ID 网络以建立用于槽位填充和意图识别任务的相互关联的机制. Zhang 等人<sup>[39]</sup>提出了一种层次化的胶囊神经网络, 以对单词、槽位和意图之间的层次关系进行显式建模.

总的来说, 深度学习模型已经在口语理解任务上得到了充分的应用. 以上研究表明, 性能最好的模型也都是采用深度学习的技术. 接下来, 本文将从口语理解涉及的深度学习技术、数据集及联合建模方法进行系统的介绍.

## 2 口语理解涉及的深度学习技术

深度学习是机器学习的分支, 已被广泛地应用到计算机视觉<sup>[40-45]</sup>、自然语言处理<sup>[46-49]</sup>、自动驾驶<sup>[50,51]</sup>等多个领域. 口语理解是对文本进行处理获取意图和槽位标签的任务, 属于自然语言处理领域, 目前, 主要采用深度学习技术对这两个任务进行联合建模. 下面, 我们总结了 4 点口语理解相关研究中所涉及的典型的深度学习技术.

- 词嵌入: 口语理解联合建模是对文本的建模, 首先要对文本进行向量化表示. 早期的表示方法是采用 One-Hot 表示, 它是一种离散的表示法, 在处理大语料时, 存在维度高、语义鸿沟的问题. 为了降低维度和表现词之间的语义关系, 人们尝试了很多方法, 比如共现矩阵<sup>[52]</sup>、LDA(latent dirichlet allocation)<sup>[53]</sup>和 LSA(latent semantic analysis)<sup>[54]</sup>. 而词嵌入(word embedding)<sup>[55-59]</sup>的出现, 可以很好地解决上述问题. 经过词嵌入层处理后, 每一个单词都被映射到一个低维、稠密的实值向量上, 向量之间可以表达更多的语义信息. 本文所介绍的文献<sup>[7-9,12-15,37,60,61]</sup>建模时, 都会引入一个嵌入表示层, 在使用反向传播<sup>[62]</sup>和梯度下降算法学习过程中, 这些嵌入向量也会自动地更新到合适的状态. 部分文献还加入实体嵌入向量和字级别嵌入向量来提升模型学习的能力.
- 循环神经网络(RNN): 词嵌入所获取的向量, 可以输入到后续的编码器层, 用于建模更高层次的句法或语义信息. 这里介绍的 RNN 结构包括普通 RNN<sup>[63]</sup>、LSTM<sup>[64]</sup>和 GRU<sup>[65]</sup>这 3 种, 它们是专门设计

用来处理序列问题的网络架构. Liu 等人<sup>[9]</sup>采用双向 LSTM 来学习句子的上下文语义信息, LSTM 的门控机制有效地解决了普通 RNN 所面临的梯度消失问题, 使其具有长距离特征捕获能力. 另外, Zhang 等人<sup>[7]</sup>使用双向 GRU 对其进行建模, 带有重置门和更新门的 GRU 结构, 一方面简化了 LSTM 的结构, 降低了网络参数量; 另一方面, 还能保持与 LSTM 相当的长序列建模能力. 它是口语理解任务建模时常用的一种网络结构.

- 注意力机制: 无论是机器学习、深度学习还是强化学习, 我们似乎都在寻找一种可以模拟人脑的机制, 注意力机制(attention mechanism)实际上也是源于此种设计思路. 注意力可以应用于任何类型的输入而无论其结构如何. 在计算能力有限情况下, 注意力机制是解决信息过载问题的主要手段, 还是一种资源分配方案, 将计算资源分配给更加重要的任务. Bahdanau 等人<sup>[10]</sup>首次将注意力机制应用到 NLP 领域, 计算方法是: 在解码时, 使用前一个时刻的隐藏向量与输入序列做关联计算, 考虑每一个输入词与当前输出词的对齐关系, 对齐越好的词, 会有越大的权重, 对生成当前输出词的影响也就越大. Chen 等人<sup>[61]</sup>提出了一种基于 BERT 模型架构的联合建模方法. 它是一种深层双向的 Transformer<sup>[66]</sup>编码器结构, 通过对大量文本数据的学习, 学到的深层次的语义信息可以很好地迁移到 SLU 任务上. Transformer 的核心就是注意力机制, 拥有强大的上下文建模能力和大规模并行计算能力.
- 胶囊网络: 在 2017 年的 NIPS 上, Sabour 等人<sup>[67]</sup>提出了胶囊网络(capsule network). 它是一个用于模仿人类视觉系统的新架构, 你可以把它想象成要用 capsule 神经元取代原来的 MP 神经元(MP neuron), 原来的 neuron 输入输出是一个标量, 而 capsule 输入是一个向量, 可以解 CNN 中组件的朝向和空间上的相对关系被忽略的问题<sup>[68-70]</sup>. Zhang 等人<sup>[39]</sup>设计了 3 种胶囊: WordCaps, IntentCaps 和 SlotCaps, 建立显式的单词、意图和槽位之间的层次关系, 结构清晰, 可解释性强. 网络更新时, 通过动态协议路由机制, 将学到的槽位信息增强意图标签的预测, 还涉及一种重路由协议机制, 将学习到的意图信息用作完善槽位标签的预测.

### 3 常用的公开数据集

目前, 常用的口语理解公开数据集主要有两个: ATIS(航空公司旅行信息系统)数据集和 Snips 数据集.

- ATIS 数据集<sup>[71]</sup>被广泛用于 SLU 研究. 数据集包含预定航班的人员的录音, 训练集包含 4 478 条话语, 开发集包含 500 条话语, 测试集包含 893 条话语. 训练集中有 120 种槽位标签和 21 种意图类型.
- Snips 数据集<sup>[72]</sup>是从 Snips 个人语音助手收集来的. 该数据集每个意图的样本数量大致相同. 训练集包含 13 084 条话语, 开发集包含 700 条话语, 测试集包含 700 条话语. 训练集中有 72 种槽位标签和 7 种意图类型.

与单域的 ATIS 数据集相比, Snips 更为复杂, 这主要是由于意图的多样性和更大的词汇量. 表 1 中显示了意图和相关的的话语示例. 关于意图多样性, 例如 Snips 中的 GetWeather 和 BookRestaurant 来自不同的意图, 词汇量较大. 另一方面, ATIS 中的意图全部是关于航班信息, 它们之间有相似的词汇. 此外, ATIS 的意图的样本数量高度不平衡, 其中, atis\_flight 大约占训练数据的 74%, 而 atis\_cheapest 仅出现一次. 表 2 中给出了这两个数据集的比较与下载地址.

除了上述两个广泛应用的基准数据集外, 还有 4 个不常用且未公开的数据集. 按照领域分类, 我们将其归纳为两类: 单域数据集和多域数据集. 单域数据集有 1 种: Car navigation, 在 Zhu 等人<sup>[35]</sup>的文献中被提到. 它是一种用于汽车导航领域的自定义的中文数据集, 训练集包含 8 000 条话语, 开发集包含 2 000 条, 测试集包含 1 944 条. 文本经过 ASR 输出后的单词错误率为 4.75% 和句子错误率为 23.42%. 而多域数据集有 3 种: 4 domains, 3 domains 和 CQUD. 其中, CQUD 是 Zhang 等人<sup>[7]</sup>从百度知道(<http://zhidao.baidu.com/>)平台收集而来的 3 286 条样本的中文数据集. 文献中并未给出具体的数据集划分方式. 详细的统计信息见表 3.

表 1 Snips 数据集中的意图和话语样本

Intent	Utterance example
SearchCreativeWork	Find me the I, Robot television show
GetWeather	Is it windy in Boston, MA right now?
BookRestaurant	I want to book a highly rated restaurant tomorrow night
PlayMusic	Play the last track from Beyonc off Spotify
AddToPlaylist	Add Diamonds to my roadtrip playlist
RateBook	Give 6 stars to Of Mice and Men
SearchScreeningEvent	Check the showtimes for Wonder Woman in Paris

表 2 ATIS 和 Snips 数据集的统计信息

	ATIS	Snips
Vocabulary size	722	11 241
Average sentence length	11.28	9.05
#Slots	120	72
#Intents	21	7
Training set size	4 478	13 084
Development set size	500	700
Testing set size	893	700
Download address	<a href="https://www.kaggle.com/weipengfei/atis-snips">https://www.kaggle.com/weipengfei/atis-snips</a>	

表 3 其他数据集的统计信息

Data set	Vocabulary size	Training set size	Development set size	Testing set size	#Slots	#Intents	
Car navigation	–	8 000	2 000	1 944	–	–	
4 domains	Alarm	433	8 096	1 057	846	8	16
	Calendar	1 832	21 695	3 626	2 555	18	20
	Communication	4 336	13 779	2 662	1 529	20	25
	Technical	2 180	7 687	993	867	18	5
3 domains	Movie	–	685	98	196	14	3
	Food	–	688	98	197	15	3
	Home	–	482	69	138	16	3
CQUD	–	–	–	–	20	43	

#### 4 口语理解联合建模

如果一个用户向系统发出请求“查找周星驰最近的喜剧”，系统能够自动识别用户以自然语言表达的意图（查询电影），并且提取实现目标的相关依据或内容（周星驰、喜剧）。通常，这两个任务都是同时出现的，彼此之间也有相互的关联。因此，对意图识别任务和槽位提取任务进行联合建模是可行的。一些研究者也关注到了这个问题，开始使用联合建模方法，实验结果表明，性能比分离任务有了较大提升<sup>[7,9]</sup>。联合建模可以只训练一个模型，大大简化单独建模的意图识别和槽位填充模型，可以降低管道所带来的错误传递，提升任务性能。本文对相关文献进行了整理和总结，将基于深度学习的口语理解联合建模算法根据模型功能分为无关联的联合建模、意图关联槽位的联合建模和双向关联的联合建模，如图 3 所示，每类方法根据实现技术特点的不同又分为多种类别，本节将详细阐述和分析这些模型方法。

为了清晰起见，表 4 中列出了本文使用的符号及其描述。

表 4 符号和描述

符号	描述
$X$	输入序列
$Y$	输出序列
$M$	意图标签总数
$\mathcal{Y}(x)$	槽位标签的搜索空间
$y^I$	意图标签的预测分布
$y^S$	槽位标签的预测分布
$d_k, i_k, \langle EOS \rangle$	域标签与意图标签组合在一起的新标签、附加标记
$W^l, W^s, W^f, W^b,$ $W_{RR}$	权重

表 4 符号和描述(续)

符号	描述
$\vec{b}^S, \vec{b}^I$	偏置
$\vec{h}_t, \vec{h}_t^I, \vec{h}_t^S$	$t$ 时刻前向隐藏向量、后向隐藏向量、前后向级联的隐藏向量
$c_t, c_t^I, c_t^S, v^I$	$t$ 时刻注意力向量、意图注意力向量、槽位注意力向量、意图嵌入向量
$H$	每个时刻产生的隐藏向量组成的矩阵
$f_{att}(H)$	注意力函数
$r_t^I, r_t^S$	意图增强向量、 $t$ 时刻槽位增强向量
$A_{y_k^S \rightarrow y_{k+1}^S}, P_{k, y_k^S}$	第 $k$ 个词-第 $k+1$ 个词的标签转移得分、第 $k$ 个词标签为 $y_k^S$ 的得分
$\odot, \oplus$	按元素相乘、按元素相加
$g(Y_i, Z)$	建模意图 $Z$ 和槽位 $Y_i$ 之间关联关系的特征函数
$\phi(\cdot), \psi(\cdot)$	LSTM 网络函数
$concat(\cdot, \cdot)$	向量级联
$v_k, u_l$	槽胶囊、意图胶囊分别通过路由协议算法得到的表示向量
$p_{kit}, q_{ilk}$	$t$ 时刻的单词被识别为第 $k$ 个槽标签的表示向量、第 $k$ 个槽标签对应于第 $l$ 个意图标签的表示向量

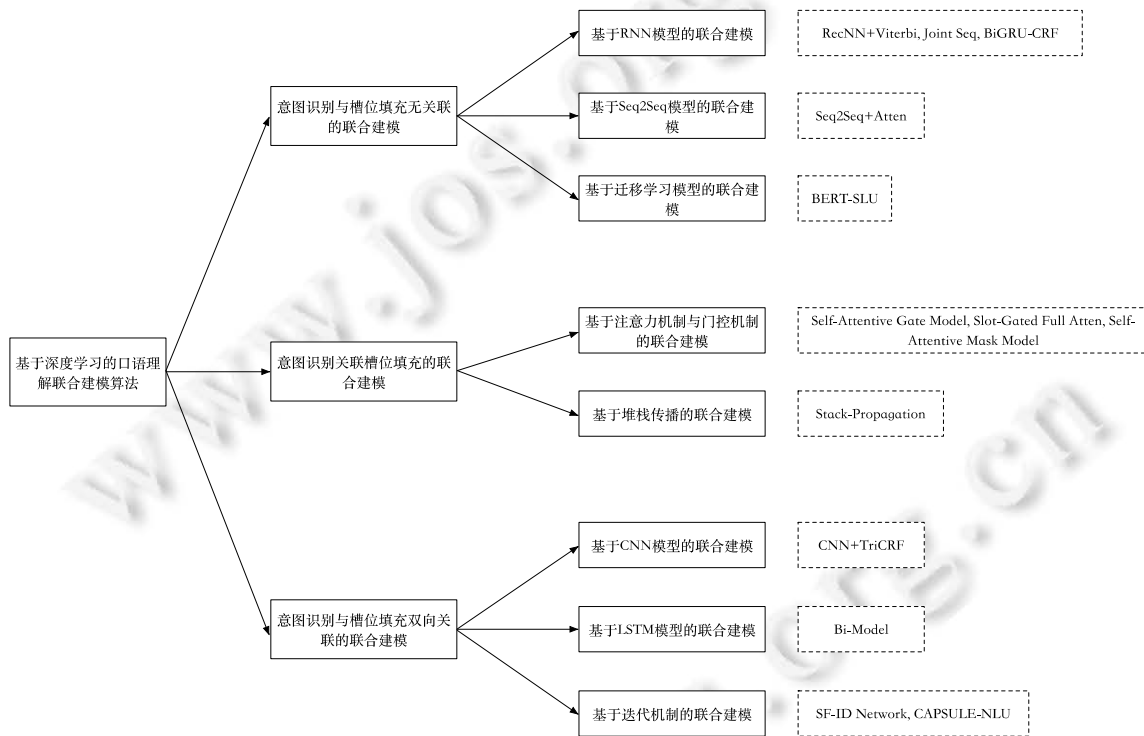


图 3 基于深度学习的口语理解联合建模算法

4.1 意图识别与槽位填充无关联的联合建模

一段时间, 研究者们将视野集中在意图和槽位的特征提取、共享表示的联合建模算法研究中. 它能够解决经典方法中存在的手工特征选择、特征提取、特征映射这 3 大主要问题. 神经网络自动学习底层特征的组合形式, 产生更加抽象的高层特征表示. 研究主要集中在编码层, 设计模型更好地去学习两个任务的共享表示, 未考虑意图识别任务和槽位填充任务在输出层的关联关系. 下面将以无关联的联合建模展开介绍. 如图 4 所示.

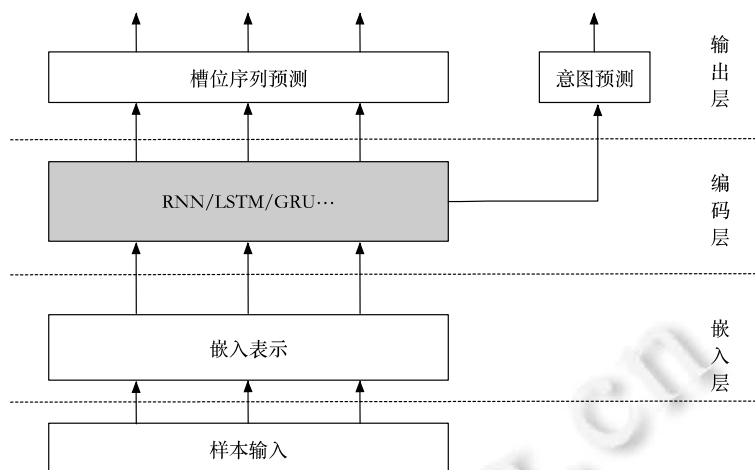


图4 意图识别与槽位填充无关联的联合建模的解决框架

#### 4.1.1 主要流程

基于意图识别与槽位填充无关联的联合建模的主要流程可以描述为：首先，针对已经标注好的训练数据集样本进行词表构建、样本数值化(单词索引)；之后，经过嵌入层输入的单词索引被转换为一个向量表示；然后，编码器层采用 RNN, LSTM, GRU 等序列建模方法学习整个序列(单词序列)的语义信息；最后，将学习到的高层序列语义信息分别用于槽位预测与意图预测。

#### 4.1.2 主流模型介绍

##### (1) 基于 RNN 模型的联合建模

与传统的多层感知机(multilayer perceptron, MLP)不同, RNN 模型隐藏层之间的节点不再是无连接的. 它不仅包含输入层的输出向量, 还包含上一时刻隐藏层的输出向量, 可以利用其内部的这种记忆块来处理任意长度的序列数据, 学习任意长度的句子的嵌入向量表示; 同时, 与全连接神经网络不同的是, RNN 模型各层之间的参数是共享的, 大大降低了网络中需要学习的参数量以及模型的复杂度.

2014 年, 微软研究院的 Guo 等人<sup>[8]</sup>首次将递归神经网络(RecNN)应用到口语对话系统中的口语理解(SLU)任务上. 递归神经网络主要用来建模自然语言句子的语义信息: 给定一个句子的语法结构(一般为树状结构), 可以使用递归神经网络, 按照语法结构来编码一个句子的语义. 此方法为每句话生成对应的一个句法树, 树中的叶子节点对应句子中的单词, 输入为单个词向量, 每个词性可以被当作权重向量. 这样, 每个词在其路径上的运算为简单的词向量与权重向量的点积运算. 父节点的值是多个分支节点与权重向量的点积和, 再经过激活函数之后的结果. 其中, 向量捕获单词组成部分的固有意义, 而父节点的运算可以捕获相邻单词或短语的含义. 对于意图识别来讲, 文中直接采用融合了句子信息的根节点的输出向量, 通过 softmax 函数计算出可能的目标意图类别. 而对于槽位填充而言, 为了包含更多的上下文信息, 文中将前一个和下一个路径向量与当前路径向量连接起来, 形成一个三路径向量(tri-path), 每个融合了上下文单词信息的 tri-path 用作多类最大熵分类器的输入, 以预测相应单词的 IOB 槽位标签. 这种 RecNN 的递归机制<sup>[73]</sup>能够比传统的 MLP 更好地捕获序列信息. 此外, 它还利用了更加丰富的句法树信息, 能够更好地建模句子的结构化信息, 并且在语言上更加合理、可解释.

Hakkani-Tür 等人<sup>[36]</sup>在 2016 年提出了一种 RNN-LSTM 架构, 对多域多任务(域和意图检测, 槽位填充)进行联合建模. 该模型支持多任务. 训练时, 来自每个域的数据可以互相补充, 简化了单一模型架构. 它在 LSTM 结构的基础上提出了两种扩展: 环视 LSTM(LSTM-LA)和双向 LSTM(bLSTM). 与 Guo 等人<sup>[8]</sup>提出的模型不同的是, Hakkani-Tür 没有使用树状结构的 RecNN, 而是使用了一种线性序列结构 RNN<sup>[74]</sup>的变体 LSTM<sup>[64]</sup>. LSTM<sup>[64]</sup>网络简化了树状结构, 消除了生成语法树所带来的时间和人工成本, 使用线性序列结构计



算成本降低, 在普通的 RNN<sup>[74]</sup>基础上增加了门控机制, 有效地解决普通循环神经网络的梯度爆炸或梯度消失问题. 文中对提出的两种扩展做了深入的分析, 这两种方法在每个时刻输入时, 都会将当前词的上下文信息建模进去, 只不过环视 LSTM 的方法只关联了前一个词和后一个词的信息; 而双向 LSTM 由于其双向的记忆能力, 可以建模更长的上下文信息来增强网络的能力. 因此, 双向 LSTM 在联合建模时比环视 LSTM 效果好, 这也说明了建模更长的上下文语义信息更加重要.

2016 年, 北京大学的 Zhang 等人<sup>[7]</sup>提出了一种基于双向 GRU(BiGRU)和最大池化的联合建模的口语理解模型. GRU 也是一种 RNN<sup>[74]</sup>的变体, 同时也使用门控解决了普通的 RNN 无法捕获长序列信息的能力. 与 Guo 等人<sup>[8]</sup>提出的模型相比, 本文的模型不采用树状结构, 结构简单, 计算效率提升. 与 Hakkani-Tür 等人<sup>[36]</sup>的模型相比, GRU 采用两个门控(更新门、重置门), 简化了 LSTM 的三门控(遗忘门、输入门、输出门)结构, 可以等效地获取长序列建模能力的同时, 又简化了 LSTM 的参数数量. 图 5 描述了用于口语理解无关联的联合建模的体系结构. 图 5 最下部分, 网络对输入句子提取多个级别的特征向量. 它主要包括两个部分: 单词嵌入向量和命名实体嵌入向量. 图 5 中间部分, 编码器层采用了双向 GRU, 除了学习句子从前到后的信息之外, 还可以学习到句子从后向前的信息, 大大提高了建模上下文信息的能力. 图 5 右侧部分和最上部分的输出层分别是意图预测与槽位预测: 意图预测采用了最大池化, 用于捕获句子的全局特征, 有利于意图识别; 槽位预测增加了条件随机场(CRF)模拟标签转移, 以此获取整个序列的全局最优解.

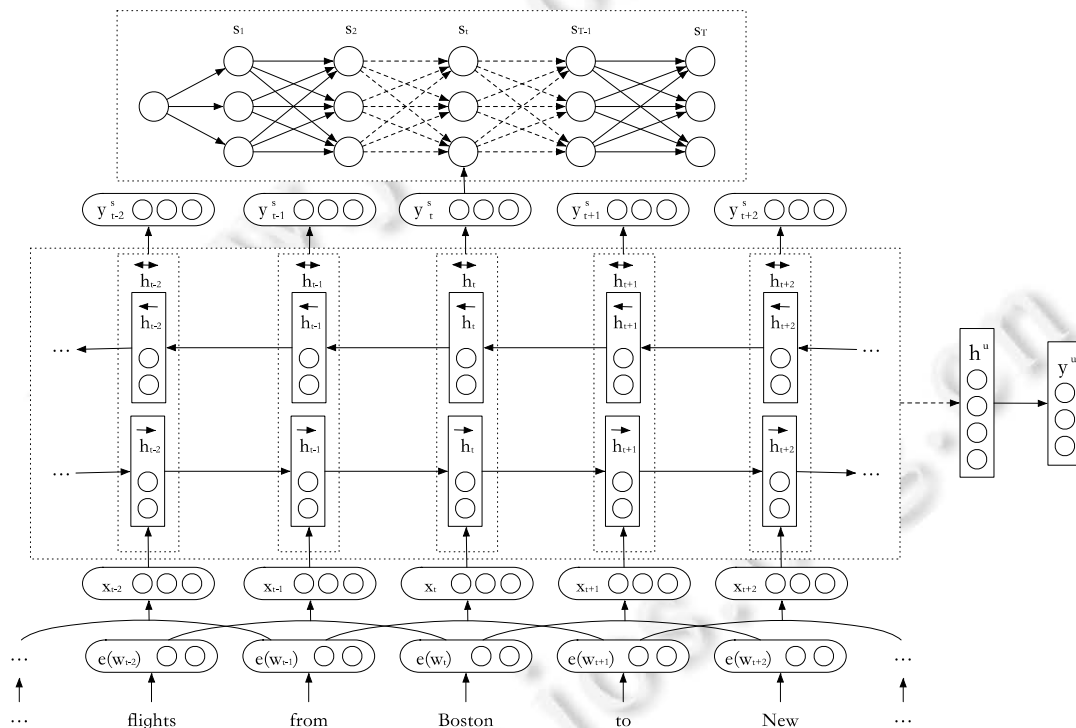


图 5 双向 GRU 和最大池化

## (2) 基于 Seq2Seq 模型的联合建模

随着序列到序列的神经网络在监督学习上的应用<sup>[75,76]</sup>, 基于注意力机制的编解码架构的神经网络模型<sup>[77]</sup>近年来在机器翻译<sup>[78]</sup>和语音识别<sup>[79]</sup>中取得了令人鼓舞的结果. 受此启发, 卡内基梅隆大学的 Liu 等人<sup>[9]</sup>在 2016 年提出一种基于注意力的编解码模型, 首次应用到口语理解的联合建模任务上. 这种架构结构清晰, 通常包括编码器部分和解码器部分. 图 6 左边部分是接收输入的编码器部分, 这部分也是采用能够更好地建模序列上下文信息的双向 LSTM (BiLSTM)结构. 图 6 右边部分是解码器输出部分, 右上部分用于意图预测, 右

下部分用于槽位预测. 由于槽位预测是一种序列标注任务, 可以看到, 槽位的输出序列与编解码模型的输入序列是等长的. 与基于 RNN 模型的联合建模中的 3 种模型相比, 编解码架构新颖, 任务可扩展性强, 输入输出序列还可以是变长的. 与 Zhang 等人<sup>[7]</sup>的模型相比, 这里使用 BiLSTM, 与 BiGRU 有等效的长距离序列上下文建模能力, 没有 CRF 层, 消除了 CRF 带来的计算复杂度高、时间效率低的问题. 加入了注意力机制, 可以学习编解码的对齐信息, 关注重要的信息, 对于序列标注任务更加有效.

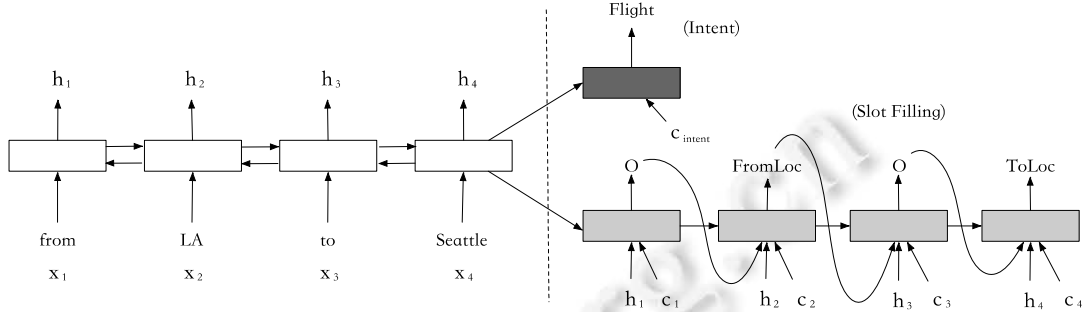


图 6 带有对齐的输入和注意力的编解码模型

(3) 基于迁移学习模型的联合建模

口语理解任务上也经常受到人工标注的小规模训练数据的影响, 导致泛化能力较差, 特别是对稀有单词的泛化能力更差. 为了解决这种问题, 阿里巴巴语音实验室的 Chen 等人<sup>[61]</sup>提出了一种基于 BERT<sup>[80]</sup>的联合意图识别和槽位填充模型, BERT 的网络架构是一种深层双向的 Transformer 编码器<sup>[66]</sup>结构. 图 7 最下部分是句子输入, 与基于 RNN 和 Seq2Seq 的模型都不同, 主要是: (1) 采用 WordPiece<sup>[81]</sup>分隔算法; (2) 句子首位插入两个特殊标记, 句首是分类嵌入[CLS], 句尾是[SEP]. 其中, WordPiece 比传统空格分隔 tokenization 技术能够更好地处理未知或罕见的词汇(OOV)问题, 而特殊标记[CLS]作为分类标识, 可以将意图识别任务与槽位序列标记任务联合起来. 图 7 中间部分采用自注意力机制(self-attention), 与 RNN 形式的模型不同, self-attention 机制可以不考虑单词与单词之间的距离, 而直接计算相互的依赖关系, 能够很好地学习到一个句子的内部结构, 还可以实现并行计算, 解决 RNN 无法并行、计算耗时的问题. 与 Liu 等人<sup>[9]</sup>的模型类似之处是都采用了注意力机制, 不同之处在于注意力机制中查询(Q)的不同: Liu 等人<sup>[9]</sup>因其结构特性, 使用的查询(Q)来自于解码器, 键(K)和值(V)都来自于编码器层; 而 Chen 等人<sup>[61]</sup>的模型中查询(Q)、键(K)和值(V)都出自编码器. 虽然向量上有所不同, 但是建模序列能力上是等效的.

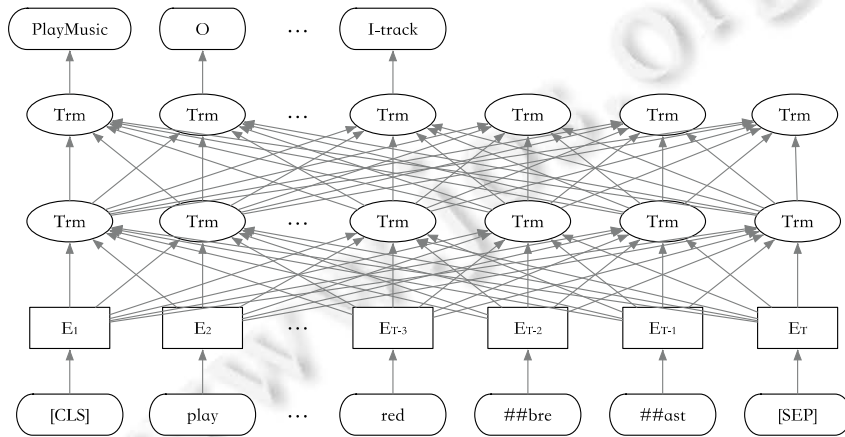


图 7 基于 BERT 的联合槽位填充和意图检测

4.1.3 无关联的联合建模存在的共性问题

然而, 无关联的联合建模存在以下几个缺点.

- (1) OOV 问题: 我们一般把数据集分成训练集、开发集和测试集, 通常采用训练集的数据构建词表, 因此除 BERT 以外, 其他模型都采用传统的空格分隔的 tokenization 技术, 无法处理未知或罕见的词汇问题.
- (2) 并行计算: 虽然 RNN 相比于 MLP 网络更适合处理序列数据问题, 但是 RNN 及其变体 LSTM, GRU 在处理内部计算时仍然存在循环结构, 难以并行化.
- (3) 忽视了两个任务之间的联系: 口语理解任务包含意图识别和槽位预测. 槽位预测结果产生的实体约束往往是完成某项任务意图所需要的, 这里的无关联建模没有考虑到这种问题.

4.1.4 相关文献整理及核心公式

意图识别和槽位填充无关联的联合建模相关文献整理及核心公式见表 5.

表 5 意图识别和槽位填充无关联的联合建模模型及核心公式

文献	年份	模型	核心公式	公式类型
Guo 等人 <sup>[8]</sup>	2014	递归神经网络、维特比算法(RecNN+Viterbi <sup>[82]</sup> )	$y^I = \arg \max_{i \in M} P(y_i^I   x); y^S = \arg \max_{y^S \in \mathcal{Y}(x)} P(y^S   x)$	分类
Hakkani-Tür 等人 <sup>[36]</sup>	2016	双向 LSTM、环视 LSTM(Joint Seq)	$X=x_1, \dots, x_t, \langle EOS \rangle; Y=s_1, \dots, s_t, d_k, i_k;$ $p_t = \text{softmax}(W^f \bar{h}_t + W^b \bar{h}_t)$	分类
Zhang 等人 <sup>[7]</sup>	2016	双向 GRU、最大池化 (BiGRU-CRF)	$\bar{h}_t = \overline{GRU}(x_t, \bar{h}_{t-1}); \bar{h}_t = \overline{GRU}(x_t, \bar{h}_{t+1});$ $\bar{h}_t = [\bar{h}_t, \bar{h}_t]; y_t^s = \text{softmax}(W^s \bar{h}_t + b^s);$ $h^I = \max_{t=1}^T \bar{h}_t; y^I = \text{softmax}(W^I h^I + b^I)$	分类
Liu 等人 <sup>[9]</sup>	2016	编解码架构、注意力机制(attention BiRNN)	$\hat{y} = \arg \max_y P(y   x);$ $P(y) = \prod_{t=1}^T P(y_t   y_{t-1}, h_{t-1}, c_t)$	分类、 条件概率
Chen 等人 <sup>[61]</sup>	2019	BERT(BERT SLU)	$p(y^I, y^S   x) = p(y^I   x) \prod_{t=1}^T p(y_t^S   x)$	联合概率

4.2 意图识别关联槽位填充的联合建模

从早期的无关联建模的研究中可以看到: 双向 LSTM/GRU 在句子建模上有比较大的优势, 所涉及的模型也都具有一定的竞争优势. 而注意力机制最早作为基于 RNN 的编解码框架的一部分被 Bahdanau 等人<sup>[10]</sup>提出, 首次应用在机器翻译模型中, 获得了当时 SOTA (state-of-the-art) 的结果. 随后, 一些研究者们对这两种网络在 SLU 任务上做了进一步探索. 与编解码架构中注意力机制不同之处是, 这里采用自注意力机制学习单词之间的相互依赖关系和上下文信息, 同时也考虑了两个任务之间的单向关联关系. 下面将以意图关联槽位的联合建模展开介绍, 如图 8 所示.

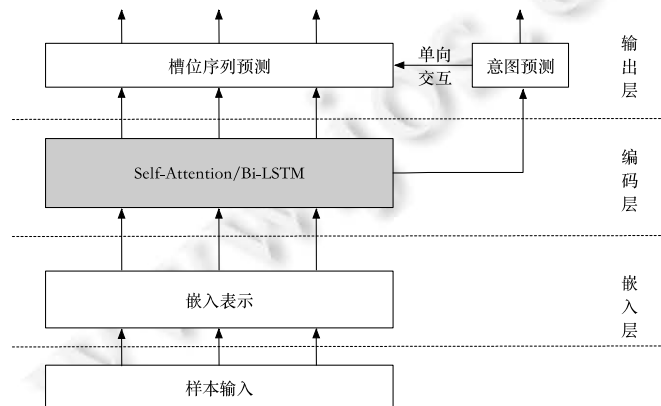


图 8 意图识别关联槽位填充的联合建模的解决框架

#### 4.2.1 主要流程

基于意图关联槽位的联合建模与第 4.1 节提到的无关联联合建模相比有两处不同: (1) 编码器层加入了新的网络结构 **self-attention**, 自注意力机制除了在学习长序列上下文依赖关系上比较好之外, 还可以在计算上实现并行, 时间效率高; (2) 输出层槽位序列预测时融入了意图预测信息, 意图信息能为槽标签提供更多目标性的信息, 这里采用门控机制进行两个任务之间的单向交互。

#### 4.2.2 主流模型介绍

##### (1) 基于注意力机制和门控机制的联合建模

针对无关联模型中存在的忽略了两个目标任务之间关联关系的问题, 金山 AI 实验室和清华大学的 Li 等人<sup>[13]</sup>提出了一种基于门控的自注意力模型, 首次将意图信息用于槽位预测, 以此充分利用意图和槽位之间的语义相关性, 其模型如图 9 所示. 在该模型中, **self-attention** 层的输出都与输入有个 **concat**, 这种网络结构叫做残差连接<sup>[70]</sup>, 可以解决深层次网络退化问题, 也可以降低一定的信息丢失问题. 与 Zhang 等人<sup>[7]</sup>的模型不同, 词嵌入层将命名实体嵌入(named entity embedding)换成了字符级嵌入(character embedding)<sup>[83,84]</sup>, 在解决 OOV 问题上更有效. 编码器层使用了两个 **self-attention**, 其中, 第 1 个 **self-attention** 更多地去捕获长序列长距离词之间的依赖关系, 用于共享表示学习; 第 2 个 **self-attention** 将学习到的词依赖关系用于槽位序列标签的预测, 因为对于序列标注来讲, 更多的是需要 **word-level** 的信息; 而对于意图分类预测, 更多地需要 **sentence-level** 的信息. 门控(**gate**)的引入, 为第 4.1 节无关联建模的模型加入了任务之间的关联, **gate** 的设计最初来源于 LSTM, 用于解决普通 RNN 建模长序列时所带来的梯度消失问题. **gate** 更像是一种信息选择器, 它使用意图(**intent**)预测信息将句子中有实体约束的单词选出来, 有利于槽位序列标签的预测. 最后, 模型通过端到端联合训练, 同时优化两个任务的目标.

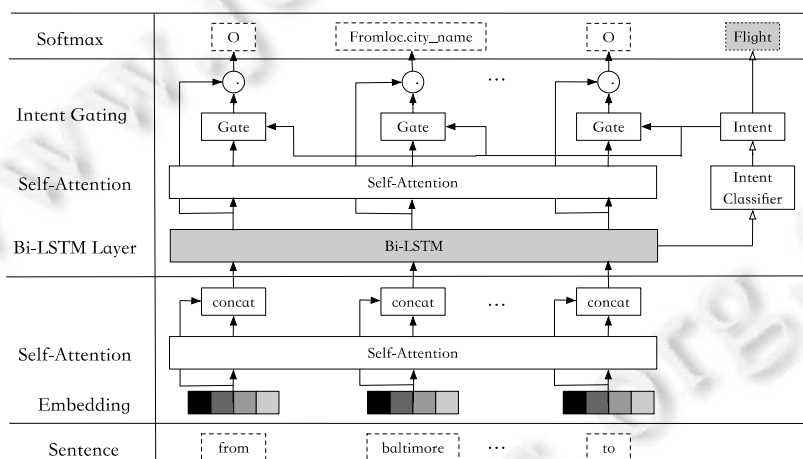


图 9 具有门控机制的自注意力模型

台湾大学信息研究所的 Goo 等人<sup>[12]</sup>提出了一种基于意图注意力、槽位注意力和门控机制模型. 与 Li 等人<sup>[13]</sup>的模型类似, 都采用了 2 个注意力层和 1 个 BiLSTM 层: 一个是意图注意力(**intent attention**)层, 被设计用于学习句子中意图特征; 另一个是槽位注意力层, 专注于学习槽位序列特征. 对于门控机制而言, 同样是将意图预测信息用于槽位序列预测上, 有利于相关联的实体约束单词的识别. 在词嵌入层, 本文并没有加入额外的命名实体嵌入或字符级嵌入, 但是与 Li 等人<sup>[13]</sup>的模型相比, 性能上并没有明显的下降. 这表明了双向 LSTM (BiLSTM)和自注意力机制(**self-attention**)在句子序列建模和上下文特征学习上十分有效, 而意图任务对于槽位填充任务也有帮助, 只是名字上的命名不同, Goo 等人<sup>[12]</sup>将其命名为 Slot Gate.

随着 Transformer 模型<sup>[66]</sup>的提出, 越来越多的人受到其出色的性能表现的吸引, 开始挖掘其内部结构的可迁移性. 字节跳动公司的 Chen 等人<sup>[14]</sup>于 2019 年利用其内部的多头注意力机制, 提出了一种用于解决 SLU 任

务的联合模型. 对于嵌入层来说, 与 Zhang 等人<sup>[7]</sup>的模型不同, 与 Li 等人<sup>[13]</sup>的模型相同, 都在原先词嵌入向量的基础上加入了字符级嵌入向量, 可以缓解存在的 OOV 问题. 对于编码器层来讲, 都采用了注意力机制和 BiLSTM, 不同之处在于: 本文使用了一种多头注意力机制(multi-head local self-attention), 一个 self-attention 可以学到单词依赖之间的单方面的程度差异, 从而将模型分成多个头, 形成多个子空间, 可以使模型去关注不同方面类型的差异; 同时, 多头注意力层可以进行并行计算, 大大降低了时间成本. 与 Li 等人<sup>[13]</sup>的模型相同, 这里也使用 BiLSTM 的隐藏层输出用于意图识别的预测. 对于输出层, 本文将意图预测信息用于槽位填充任务中, 增强槽位信息. 我们者在文中将其命名为 Mask 机制, 从形式和运算上来看, 与 Li 等人<sup>[13]</sup>和 Goo 等人<sup>[12]</sup>的模型中提到的门控机制是一样的功能. 最后, 使用了 CRF 层约束槽位填充(SF)的输出, 获取全局最优的序列标签. 此模型被应用于在线教育平台, 提供一定的情景对话, 提升中小学生学习英语口语和听写能力.

### (2) 基于堆栈传播的联合建模

门控机制也有一定的不足, 它不能充分地总结和记住意图信息, 并且在槽位填充和意图识别的交互是通过隐藏向量来交互, 造成了整个框架可解释性低. 除了上述门控机制用作这种任务间的单向交互之外, 哈尔滨工业大学 SCIR (research center for social computing and information retrieval)的 Qin 等人<sup>[37]</sup>提出了一种新颖的堆栈传播(stack-propagation)框架, 引入投票机制整合意图信息, 从而进一步指导槽位填充任务. 图 10 最下部分 Encoder 层使用了之前模型中已经验证过高效的 self-attention 和 BiLSTM 结构, 其编码输出向量信息中不仅包含词级别的信息, 还包含句子级别的信息. 图 10 中间部分的 Token-level Intent Detection decoder 层与 Goo 等人<sup>[12]</sup>、Li 等人<sup>[13]</sup>和 Chen 等人<sup>[14]</sup>所采用的门控机制不同: 一方面, 采用 stack-propagation 框架直接将意图识别的输出信息作为槽位填充任务的输入, 以指导槽位填充任务的预测, 提高了模型的可解释性; 另一方面, 创新性地提出一种单词级别的意图识别机制, 对每一个单词分别进行意图识别, 最终整个句子的意图由每个单词的意图结果进行统一的投票决定. 然后, 将进行单词级别的意图识别信息传递给对应单词的槽位预测节点, 可以缓解一定的错误传播问题.

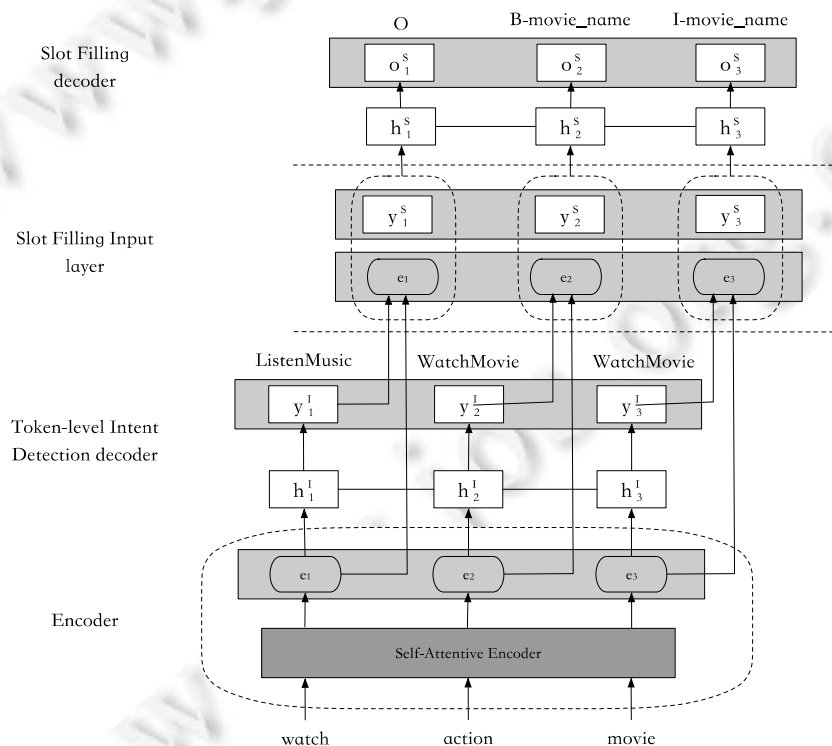


图 10 联合意图识别和槽位填充的堆栈传播框架

4.2.3 意图关联槽位的联合建模存在的共性问题

意图关联槽位的联合建模与无关联的联合建模相比, 在模型上采用了 BiLSTM、注意力机制, 能够更好地学习到长序列长距离句子级别的信息和上下文关注信息. 注意力机制可以实现并行化, 而 BiLSTM 因其内部的循环结构无法实现并行计算. 在编码阶段, 建模单词、槽位与意图信息时层次不够清晰, 可解释性较弱. 而且这种意图关联槽位的单向交互模型忽略了槽位对意图的作用, 通常, 槽位预测信息对意图识别任务也是非常重要的.

4.2.4 相关文献整理及核心公式

意图识别关联槽位填充的联合建模相关文献整理及核心公式见表 6.

表 6 意图识别关联槽位填充的联合建模模型及核心公式

文献	年份	模型	核心公式	公式类型
Li 等人 <sup>[13]</sup>	2018	自注意力、门控机制 (self-attentive gate model)	$s_t = \text{Self-Attention}(h_t, H); h_t^* = \text{MLP}([s_t, v^t]);$ $h^l = f_{att}(H); y^l = \text{softmax}(W^l h^l + b^l);$ $o_t = h_t \odot h_t^*; y_t^s = \text{softmax}(W^s o_t + b^s)$	槽位门控、分类
Goo 等人 <sup>[12]</sup>	2018	意图槽位注意力、门控机制(slot-gated full attention)	$h_t = [\bar{h}_t, \tilde{h}_t]; g = \sum v \cdot \tanh(c_t^s + W \cdot c^l);$ $y_t^l = \text{softmax}(W^l (h_t + c^l));$ $y_t^s = \text{softmax}(W^s (h_t + c_t^s \cdot g))$	槽位门控、分类
Chen 等人 <sup>[14]</sup>	2019	多头注意力、掩码机制 (self-attentive mask model)	$y^l = \text{softmax}(W^l h_n + b^l);$ $s(X, Y^s) = \sum_{k=0}^n A_{y_k^s, y_{k+1}^s} + \sum_{k=1}^n P_{k, y_k^s}$	分类、得分函数
Qin 等人 <sup>[37]</sup>	2019	堆栈传播 (stack-propagation)	$h_t^l = f(h_{t-1}^l, y_{t-1}^l, e_t); y_t^l = \text{softmax}(W_h^l h_t^l);$ $h_t^s = f(h_{t-1}^s, y_{t-1}^s, y_t^l \oplus e_t); y_t^s = \text{softmax}(W_h^s h_t^s)$	分类

4.3 意图识别与槽位填充双向关联的联合建模

意图关联槽位的联合模型一方面使用建模长序列能力更加优异的 BiLSTM 和 Attention 网络模块, 另一方面还对两个任务之间的关系进行了探索, 将意图预测信息用于槽位填充任务中, 获得了任务整体性能的提升, 在多任务学习上也做出了贡献. 同时, 许多研究人员发现: 意图识别和槽位填充这两个任务并不是相互独立的, 槽位填充任务高度依赖于意图识别的结果, 意图识别也可以指导槽位标签的预测. 一些研究者在构建 SLU 模型时考虑了意图与槽位的双向交互关系, 并且探索新的网络结构. 下面将以意图识别与槽位填充双向关联的联合建模的方法展开介绍, 如图 11 所示.

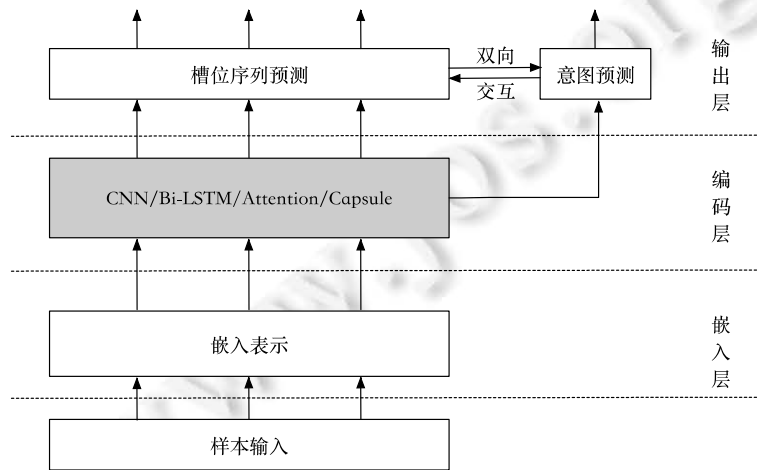


图 11 意图识别与槽位填充双向关联的联合建模的解决框架

### 4.3.1 主流流程

基于意图识别与槽位填充双向关联的联合建模与第 4.1 节、第 4.2 节所介绍的模型也有两处差异。

- (1) 编码层除了采用建模长序列上有优势的 BiLSTM/Attention 之外, 还探索了 CNN 和新颖的网络模型 Capsule<sup>[67]</sup>. CNN 在视觉领域大放异彩, 主要处理图像问题, 这里被用来建模文本序列问题. Capsule 使用动态路由机制进行更新, 网络层次结构清晰, 激活向量更具有解释性.
- (2) 从图 11 的输出层可以看到: 在意图预测和槽位序列预测之间有双向交互, 探索两个任务之间的促进关系.

### 4.3.2 主流模型介绍

#### (1) 基于 CNN 模型的联合建模

Xu 等人<sup>[60]</sup>在 2013 年提出了基于卷积神经网络和三角链条件随机场(CNN+TriCRF)的联合模型. 这个模型是最早提出用神经网络解决 SLU 任务的, 所提出的模型结构可以看作三角链 CRF<sup>[85]</sup> (TriCRF) 的神经网络版本. 图 12 中间编码部分采用 CNN 网络, CNN 可以借助多个不同尺度的卷积核提取序列特征, 其输出直接被用来做意图和槽位任务的预测. 图 12 最左上部分是槽位预测, 直接采用卷积层得到的高维特征作为输入. 图 12 最右上部分是意图预测, 对卷积层输出直接采用 max pooling 得到的整句表示向量作为输入. 最后, 输出层采用 Tri-CRF 构造特征函数建模意图与槽位之间的双向交互. Tri-CRF 模型与线性 CRF 的区别是输入之前用一个前向网络来得到每个标签的分类.

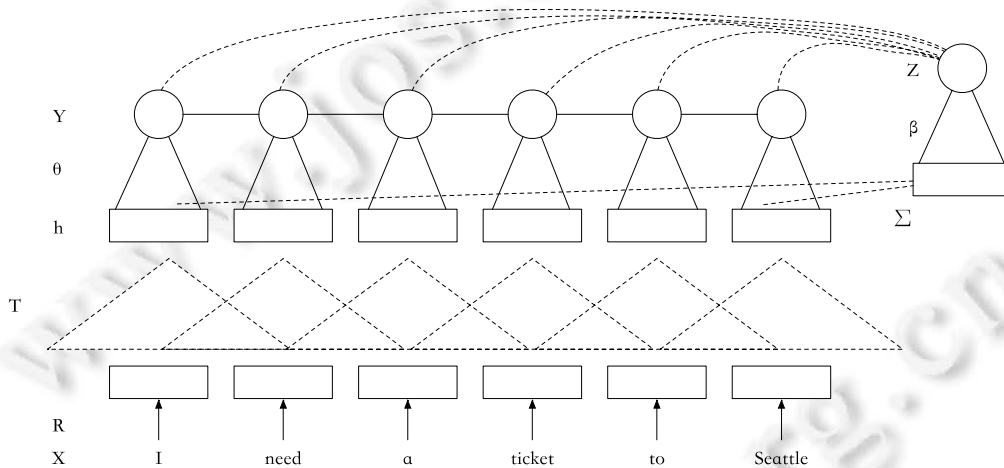


图 12 基于 CNN 的联合槽位填充和意图检测

#### (2) 基于 LSTM 模型的联合建模

与 Liu 等人<sup>[9]</sup>提出的编解码模型不同, 2018 年, 三星美国研究院的 Wang 等人<sup>[38]</sup>对输入序列分别应用 BiLSTM, 提出了一种带有解码器的双模型架构. 从图 13 中可以看到, 有上、下两个编解码层: 意图识别任务网络和槽位填充模型任务网络.

- 对于意图识别网络部分, 首先采用双向 LSTM 学习长序列句子信息, 然后将每个时刻产生的双向隐藏向量输入给一个单向 LSTM, 最后的隐藏层输出用作意图预测.
- 与意图识别部分相同, 槽位填充网络将单向的 LSTM 每个时刻产生的隐藏输出向量用于槽序列标签的预测.

在任务交互方面, 第 4.2 节采用门控机制将意图信息用于槽位填充任务的预测, Xu 等人<sup>[60]</sup>的模型是采用 Tri-CRF 构建意图和槽位的双向交互. 本文与 Qin 等人<sup>[37]</sup>的模型使用的任务交互方式相同, 都是将输出的隐藏向量直接用于另一任务的预测. 这种方式可以使意图与槽位任务之间的交互更加充分, 提高了模型的可解释性.

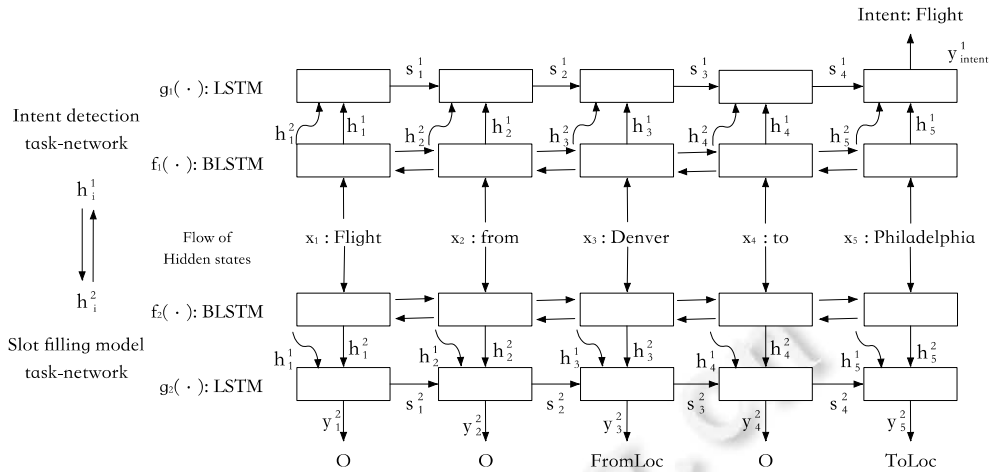


图 13 带有解码器的双模型架构

(3) 基于迭代机制的联合建模

2019 年, 北京邮电大学的 E 等人<sup>[15]</sup>使用迭代机制建模意图识别和槽位填充的双向关联关系. 图 14 是该模型的整体架构, 输入与大多模型类似, 仅采用单词嵌入, 编码器层与第 4.2 节模型所采用的网络模块相同, BiLSTM 和 self-attention 组合使用, 不仅获取了长序列句子级别的语义信息, 也得到了单词级别的上下文关注信息. 与 Goo 等人<sup>[12]</sup>的模型十分相似, 都设计了意图注意力模块和槽位注意力模块. 不同点是两个任务的交互上, 如图 14 右侧部分, 我们设计了 SF-ID 网络——ID 子网和 SF 子网来建立两个任务的信息交互, 以帮助彼此相互促进. SF 子网引入相关因子, 将意图信息应用于槽位填充任务; 而 ID 子网也计算一个相关因子, 将槽位信息用于意图识别任务. 这里的相关因子其实与门控机制相似. 然后, 这两个子网通过迭代机制进行训练. 最后, 输出层与 Zhang 等人<sup>[7]</sup>的模型相同, 在 SF 子网的输出上添加了 CRF 层, 可以获得话术标签序列的最佳链.

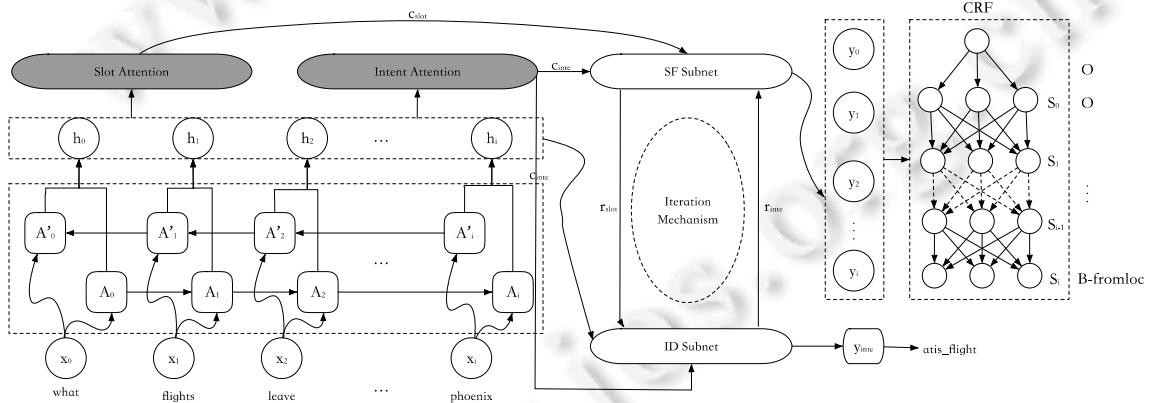


图 14 基于 SF-ID 网络的模型结构

同年, Zhang 等人<sup>[39]</sup>提出了一种新颖的基于胶囊网络的联合意图识别和槽位填充的双向关联模型. 胶囊网络是 Geoffrey Hinton 在 2017 年提出的一种新型神经网络结构, 主要是为了解决卷积神经网络的一些缺点 (例如平移不变性、大量数据来泛化等). 和 E 等人<sup>[15]</sup>的模型一样, 采用了 BiLSTM 生成每个单词的向量表示. 而与之提到的模型不同之处是: 我们将胶囊网络应用在 SLU 任务中, 建立显式的单词、槽位和意图之间的层次关系, 结构清晰, 可解释性更强. 网络更新时, 通过动态协议路由机制将学到的槽位信息增强意图标签的预测, 还涉及了一种重路由协议机制, 将学习到的意图信息完善槽位标签的预测. 文献中使用相同的数据集



与现成的系统服务(api.ai, wit.ai, snips.ai 等)做了比较, 整体性能上具有一定的优势.

4.3.3 双向关联的联合建模存在的共性问题

与无关联的联合建模模型相比, 双向关联的联合建模模型使用了具有更强序列建模能力的 BiLSTM 和 self-attention 网络模块, 同时考虑了两个任务之间的双向交互. 与意图关联槽位的联合建模模型相比, 网络编码器层使用网络结构基本相同. 在任务交互上, 后者采用门控机制将意图预测信息增强槽位序列标签预测, 前者使用迭代机制/动态协议路由机制, 使得意图预测与槽位预测能够充分利用彼此的信息实现相互的促进. 虽然双向关联模型在建模能力与性能上都达到了最优, 但是仍然存在一些不足: (1) 使用了 BiLSTM, 在并行计算上有一定的限制; (2) 当任务交互时, 迭代机制增加了超参数, 动态协议路由机制计算效率不高; (3) 传统的词嵌入无法很好地处理未知或罕见的词汇(OOV 问题), 而字级别嵌入作为 OOV 的解决方法粒度过细.

4.3.4 相关文献整理及核心公式

意图识别与槽位填充双向关联的联合建模相关文献整理及核心公式见表 7.

表 7 意图识别与槽位填充双向关联的联合建模模型及核心公式

文献	年份	模型	核心公式	公式类型
Xu 等人 <sup>[60]</sup>	2013	卷积神经网络、三角链条件随机场 (CNN+TriCRF)	$P(Y, Z   X) = \frac{e^{\sum_i (u_i(Y_i, Z_i) + g(Y_i, Z_i) + \sum_j (\theta_j(Y_j) + \beta_j(Z_j)) h_j)}$	联合概率
Wang 等人 <sup>[38]</sup>	2018	两个相关的双向 LSTM(Bi-model)	$s_t^1 = \phi(S_{t-1}^1, h_{t-1}^1, h_{t-1}^2); y_t^{1,1} = \arg \max_{y_t^1} P(y_t^1   s_{t-1}^1, h_{t-1}^1, h_{t-1}^2)$ $s_t^2 = \psi(h_{t-1}^2, h_{t-1}^1, s_{t-1}^2, y_{t-1}^2); y_t^{2,2} = \arg \max_{y_t^2} P(y_t^2   h_{t-1}^1, h_{t-1}^2, s_{t-1}^2, y_{t-1}^2)$	分类
E 等人 <sup>[15]</sup>	2019	SF-ID 网络 (SF-ID network)	$f = \sum V * \tanh(c_t^s + W * c^l); r_t^s = f c_t^s;$ $r = \sum_{t=1}^T \alpha_t r_t^s; r^l = r + c^l; f = \sum V * \tanh(c_t^s + W * r^l);$ $y^l = \text{softmax}(W^l \text{concat}(h_T, r^l)); y_t^s = \text{softmax}(W^s \text{concat}(h_t, r_t^s))$	SF 子网公式、ID 子网公式、分类
Zhang 等人 <sup>[39]</sup>	2019	胶囊网络 (CAPSULE-NLU)	$p_{klr} = \sigma(W_k h_r^l + b_k); v_k = \text{dynamic\_routing}(p_{klr}, \text{iter}_{slot});$ $q_{llk} = \sigma(W_l v_k^l + b_l); u_l = \text{dynamic\_routing}(q_{llk}, \text{iter}_{intent});$ $b_{kl} \leftarrow b_{kl} + p_{klr} + v_k + \alpha \cdot p_{kl}^T W_{RR} u_l^T$	路由公式、重路由公式

5 模型对比分析

参考鄂海红<sup>[86]</sup>在《软件学报》中对不同模型的数据集与评测标准的整理方式, 本文对不同模型的模型对比整理见表 8, 数据集及其评测标准整理见表 9. 其中, Self-Attentive Gate Model<sup>[13]</sup>和 Bi-Model<sup>[38]</sup>的评测结果来源于哈尔滨工业大学 SCIR<sup>[37]</sup>对模型的重新实现.

表 8 口语理解的联合建模对比

序号		1	2	3	4	5	6	7	8	9		10	11	12	13	
模型		文献 [8]	文献 [36]	文献 [7]	文献 [9]	文献 [61]+BERT	文献 [13]	文献 [12]	文献 [14]	文献 [37]	文献 [37]+BERT	文献 [60]	文献 [38]	文献 [15]	文献 [39]	
嵌入层表示	单词	✓	✓	✓	✓	-	✓	✓	✓	✓	-	✓	✓	✓	✓	
	实体	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	
	字符	-	-	-	-	-	✓	-	✓	-	-	-	-	-	-	
	WordPiece	-	-	-	-	✓	-	-	-	-	✓	-	-	-	-	
编码器层网络模型	CNN		-	-	-	-	-	-	-	-	-	✓	-	-	-	
	RNN	RecNN	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
		LSTM	-	-	-	✓	-	-	-	-	✓	✓	-	✓	-	-
		BiLSTM	-	✓	-	✓	-	✓	✓	✓	✓	✓	-	✓	✓	✓
		BiGRU	-	-	✓	-	-	-	-	-	-	-	-	-	-	-
	Attention		-	-	-	✓	-	-	✓	-	-	-	-	-	-	-
	Self-attention		-	-	-	-	-	✓	-	-	✓	✓	-	-	✓	-
Multi-head Attention		-	-	-	-	✓	-	-	✓	-	✓	-	-	-	-	
Capsule		-	-	-	-	-	-	-	-	-	-	-	-	-	✓	

表 8 口语理解的联合建模对比(续)

序号		1	2	3	4	5	6	7	8	9		10	11	12	13
模型		文献 [8]	文献 [36]	文献 [7]	文献 [9]	文献 [61]+ BERT	文献 [13]	文献 [12]	文献 [14]	文献 [37]	文献 [37]+ BERT	文献 [60]	文献 [38]	文献 [15]	文献 [39]
输出层 任务交互	门控机制	Gate Mask	-	-	-	-	✓	✓	-	-	-	-	-	-	-
	直接隐向量		-	-	-	-	-	-	-	✓	✓	-	✓	-	-
	迭代机制	SF-ID 网络 动态路由	-	-	-	-	-	-	-	-	-	-	-	✓	-
CRF		-	-	-	-	-	-	-	✓	-	-	✓	-	✓	-

表 9 不同模型的数据集及其评测标准(%)

口语理解 联合建模 方法	序号	模型	SNIPS			ATIS			发表 年份	发表会议
			Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)		
无关联	1	RecNN+Viterbi <sup>[8]</sup>	-	-	-	93.2	95.4	-	2014	IEEE
	2	JointSeq <sup>[36]</sup>	87.3	96.9	73.2	94.3	92.6	80.7	2016	INTERSPEECH
	3	BiGRU-CRF <sup>[7]</sup>	-	-	-	95.4	98.1	-	2016	IJCAI
	4	Attention BiRNN <sup>[9]</sup>	87.8	96.7	74.1	94.2	91.1	78.9	2016	INTERSPEECH
	5	BERT SLU <sup>[61]</sup>	<b>97.0</b>	<b>98.6</b>	<b>92.8</b>	<b>96.1</b>	<b>97.5</b>	<b>88.2</b>	2019	Computer Science
意图 关联 槽位	6	Self-attentive Gate Model <sup>[13]</sup>	90.0	97.5	81.0	95.1	96.8	82.2	2018	EMNLP
	7	Slot-Gated Full Attention <sup>[12]</sup>	88.8	97.0	75.5	94.8	93.6	82.2	2018	NAACL
	8	Self-attentive Mask Model <sup>[14]</sup>	93.9	99.7	-	96.5	98.9	-	2019	Computer Science
	9	Stack-propagation <sup>[37]</sup>	94.2	98.0	86.9	95.9	96.9	86.5	2019	EMNLP
	Stack-propagation+BERT <sup>[37]</sup>	<b>97.0</b>	<b>99.0</b>	<b>92.9</b>	<b>96.1</b>	<b>97.5</b>	<b>88.6</b>			
双向 关联	10	CNN+TriCRF <sup>[60]</sup>	-	-	-	95.4	94.0	-	2013	IEEE
	11	Bi-model <sup>[38]</sup>	93.5	97.2	83.8	95.5	96.4	85.7	2018	NAACL
	12	SF-ID Network <sup>[15]</sup>	90.5	97.0	78.4	95.6	96.6	86.0	2019	ACL
	13	CAPSULE-NLU <sup>[39]</sup>	91.8	97.3	80.9	95.2	95.0	83.4	2019	ACL

表 9 中的序号与表 8 中的序号一致, 序号 1-5 是意图识别与槽位填充无关联的联合建模及其相关信息, 序号 6-9 是意图识别关联槽位填充的联合建模及其相关信息, 序号 10-13 是意图识别与槽位填充双向关联的联合建模及其相关信息. 表 9 中有 3 个评测指标: 意图识别准确率(Acc)、槽位填充 F1 得分(F1)和总体准确率(Acc). 接下来的介绍中, 我们直接使用“模型+序号”的形式代替模型名称.

• 无关联的联合建模方法

模型 1 使用 RecNN 构建网络模型, 解码时使用 Viterbi 算法. 在 ATIS 数据集上, slot 得分和 intent 得分分别获得了 93.2% 和 95.4%. 模型 1 的优点是: (1) 与传统的机器学习不同, RecNN 自动抽取序列特征; (2) 采用句法树构建的树状网络结构使用了更多的先验信息. 其缺点是: (1) 首先要生成句法树, 而句法树是在有监督的数据集上训练得到的, 时间和人工成本高; (2) 树状结构节点更新时计算复杂度高. 而在 2016 年提出的模型 2-4 都采用了一种线性链 RNN 的网络结构, 替换了模型 1 中的树状结构 RecNN, 线性链的计算效率高一些. 但是普通的 RNN 结构在训练时通常会出现长依赖的梯度消失问题, 因此, 一些研究人员设计了一种带有门控机制的 RNN 变体: LSTM 和 GRU, 有效地解决了这个问题. 模型 2 和模型 4 采用了 BiLSTM, 模型 3 使用了 BiGRU, 这两种双向的循环神经网络区别不大, 都可以很好地建模长序列的句子级别的前后向信息, 只是 BiGRU 使用两个门控(更新门、重置门)简化了 BiLSTM 的结构(记忆单元), 降低了参数量. 其中, 模型 2 与模型 4 表现相当, 在 SNIPS 数据集上也做了实验, slot 得分高于 87%, intent 得分高于 96.5%, overall 得分高于 73%. 而模型 3 在 ATIS 数据集上的表现更好, 在 slot 得分和 intent 得分上分别比模型 1 高出 2.2% 和 2.7%. 主要原因是: (1) 嵌入层表示除了词嵌入外还加入了实体嵌入表示信息, 这些实体信息对整个任务有一定的帮助作用; (2) 输出层加入了条件随机场(CRF), 模拟标签转移, 以此获取全局最优解, 常常对采用了 softmax 的序列预测结果有一定的提升. 最后, 模型 5 是 2019 年提出来的. 它是基于预训练 BERT 构建的, 将在大规模数据

集上预训练的结果迁移到 SLU 任务上. BERT 使用 WordPiece 嵌入作为输入向量, 比传统的空格分隔 tokenization 技术, 能够更好地处理未知或罕见的词汇(OOV)问题. 在网络建模上, 模型 1-4 都是基于 RNN 的模型, 虽然可以有效地抽取文本序列特征, 但是这种递归/循环结构无法实现计算上的并行. 而模型 5 采用了多头注意力机制, 不仅可以有效地捕获长序列长距离句子级或单词级上下文信息, 还可以实现 GPU 上的高速并行计算, 时间效率高. 实验结果表明: 在 ATIS 数据集上, slot 得分 96.1%, intent 得分 97.5%, overall 得分 88.2%, 比模型 1-4 性能上整体高出约为 1%-5%. 在 SNIPS 数据集上, slot 得分 97.0%, intent 得分 98.6%, overall 得分 92.8%, 比模型 1-4 性能上整体提升了大约 2%-19%.

- 意图关联槽位的联合建模与无关联联合建模方法对比

模型 6-9 对于 SLU 任务建模时, 除了使用在无关联模型中得到验证的、效果较好的 BiLSTM 外, 还融入了许多自注意力(self-attention)的使用. 注意力(attention)模块最早是作为编解码框架的一部分被 Bahdanau 等人<sup>[10]</sup>提出来, 用于解决机器翻译中解码文本序列与编码文本序列的单词之间的对齐问题. 由此, 产生出了 self-attention 模块进行特征抽取. 从实验结果上看: 模型 6-8 在 ATIS 数据集和 SNIPS 数据集上效果都很好, 比无关联模型中非迁移学习的联合建模方法的评测指标相对高一些. 其中, 模型 8 性能更好, 在 ATIS 数据集上的 intent 得分高达 98.9%, 比采用了迁移学习的模型 5 的 slot 得分和 intent 得分分别还高出 0.4% 和 1.4%. 在 SNIPS 数据集上, intent 得分比模型 5 高 1.1%, slot 得分低了 3.1%. 主要原因是: (1) 加了 attention 模块与 BiLSTM 组合使用, 建模长序列长距离句子的上下文能力更强; (2) 模型 8 使用了多头注意力机制, 在特征抽取和建模上下文信息上, 比单个 attention 模块要强, 可以学习到多种类型的关注信息; (3) 使用门控机制将意图信息用于槽位序列预测, 可以有效地缩小可预测的槽位空间. 最后, 模型 9 是由哈尔滨工业大学提出的, 在任务交互上直接通过隐藏向量来交互, 避免了门控机制不能充分的总结和记住意图信息的问题. 实验结果在 SNIPS 数据集上 slot 得分是 94.2%, 比采用了门控机制的模型表现要好. 在意图预测时, 提出了一种单词级别的意图识别机制, 通过每个单词对应的意图结果进行投票决定, 可以缓解一定的错误级联的问题. 实验结果表明, 模型 9 在两个数据集上的 overall 得分都达到了最优. 而加入了预训练的模型 9, 在实验结果上表现优异, 比基于 BERT 的模型 5 整体指标都好, 在 SNIPS 数据集上的 intent 得分更是达到了 99.0%.

- 双向关联的联合建模与意图关联槽位的建模方法对比

模型 10 是最早提出用神经网络模型对 SLU 任务进行联合建模的, 该模型编码层使用 CNN 网络提取不同尺度的序列特征, 输出层使用 Tri-CRF 对任务进行联合建模, 其中, 表 7 中定义的特征函数  $g(Y_i, Z)$  建模了意图和槽位的交互关系. 实验结果表明: 在 ATIS 数据集上, slot 得分和 intent 得分分别获得了 95.4% 和 94.0%, 比模型 6 的 slot 得分高 0.3%, intent 得分低了 2.8%. 我们将其原因归结于: (1) CRF 在序列建模上有优势, 模拟标签转移, 以此获取全局最优解; (2) CNN 网络在提取长距离特征上受到卷积感受野的限制, 比 BiLSTM 建模长依赖句子上下文信息能力弱. 模型 11-13 在建模网络模型选择上与意图关联槽位的模型一样, 都使用了 BiLSTM. 不同之处主要在任务交互上, 模型 11 与模型 9 相同都使用直接隐向量实现意图与槽位的交互, 不同之处是前者使用双向交互, 后者是单向的交互. 模型 12 与模型 13 是采用迭代机制的双向交互, 而模型 13 使用了更加新颖的胶囊网络, 构建了单词、意图和槽位上更加清晰的层次关系. 实验结果表明: 在 ATIS 数据集上的 slot 得分是 95.2%, intent 得分是 95.0%, overall 得分是 83.4%. 比模型 7 对应的指标分别高出 0.4%, 1.4% 和 1.2%. 我们将其原因归结为: (1) 意图与槽位的双向交互对任务有积极的促进作用; (2) 使用胶囊网络构建了单词、槽位和意图的清晰的层次关系, 有利于任务的学习.

## 6 总 结

口语理解任务目前是自然语言处理领域的研究热点之一, 越来越多的研究者在 EMNLP, IJCAI, NAACL 和 ACL 等重要的国际会议上提交了该领域的研究成果. 本文对口语理解任务的相关背景技术进行了阐述, 给出了任务定义、框架结构和相关的深度学习核心技术, 分别对基于深度学习的无关联的联合建模、意图关联槽位的联合建模、双向关联的联合建模方法进行了分类介绍和相关技术特点的对比分析, 列举了这些模型在

主要的公开数据集上的评测指标, 并对其在模型和实验结果上进行了详细的分析. 最后, 在表 10 中列出了主流模型的源代码链接.

表 10 意图识别与槽位填充联合建模的模型论文地址和源代码链接

序号	模型	论文链接	源代码链接
1	RecNN+Viterbi <sup>[8]</sup>	<a href="http://www.cs.cmu.edu/~zguo/publications/RecNNSLU.pdf">http://www.cs.cmu.edu/~zguo/publications/RecNNSLU.pdf</a>	-
2	JointSeq <sup>[36]</sup>	<a href="https://www.csie.ntu.edu.tw/~yvchen/doc/IS16_MultiJoint.pdf">https://www.csie.ntu.edu.tw/~yvchen/doc/IS16_MultiJoint.pdf</a>	<a href="https://github.com/yvchen/JointSLU">https://github.com/yvchen/JointSLU</a>
3	BiGRU-CRF <sup>[7]</sup>	<a href="https://www.ijcai.org/Proceedings/16/Papers/425.pdf">https://www.ijcai.org/Proceedings/16/Papers/425.pdf</a>	-
4	Attention BiRNN <sup>[9]</sup>	<a href="https://arxiv.org/pdf/1609.01454.pdf">https://arxiv.org/pdf/1609.01454.pdf</a>	<a href="https://github.com/DSKSD/RNN-for-Joint-NLU">https://github.com/DSKSD/RNN-for-Joint-NLU</a>
5	BERT SLU <sup>[61]</sup>	<a href="https://arxiv.org/pdf/1902.10909.pdf">https://arxiv.org/pdf/1902.10909.pdf</a>	<a href="https://github.com/monologg/JointBERT">https://github.com/monologg/JointBERT</a>
6	Self-attentive Gate Model <sup>[13]</sup>	<a href="https://www.aclweb.org/anthology/D18-1417.pdf">https://www.aclweb.org/anthology/D18-1417.pdf</a>	<a href="https://github.com/NinedayWang/Self-Attentive-and-Gated-SLU">https://github.com/NinedayWang/Self-Attentive-and-Gated-SLU</a>
7	Slot-gated Full Attention <sup>[12]</sup>	<a href="https://www.aclweb.org/anthology/N18-2118.pdf">https://www.aclweb.org/anthology/N18-2118.pdf</a>	<a href="https://github.com/MiuLab/SlotGated-SLU">https://github.com/MiuLab/SlotGated-SLU</a>
8	Self-attentive Mask Model <sup>[14]</sup>	<a href="https://arxiv.org/pdf/1905.11393.pdf">https://arxiv.org/pdf/1905.11393.pdf</a>	-
9	Stack-propagation <sup>[37]</sup> Stack-propagation +BERT <sup>[37]</sup>	<a href="https://arxiv.org/pdf/1909.02188.pdf">https://arxiv.org/pdf/1909.02188.pdf</a>	<a href="https://github.com/LeePleased/StackPropagation-SLU">https://github.com/LeePleased/StackPropagation-SLU</a>
10	CNN+TriCRF <sup>[60]</sup>	<a href="https://ieeexplore.ieee.org/document/6707709">https://ieeexplore.ieee.org/document/6707709</a>	-
11	Bi-model <sup>[38]</sup>	<a href="https://arxiv.org/pdf/1812.10235.pdf">https://arxiv.org/pdf/1812.10235.pdf</a>	<a href="https://github.com/ray075hl/Bi-Model-Intent-And-Slot">https://github.com/ray075hl/Bi-Model-Intent-And-Slot</a>
12	SF-ID Network <sup>[15]</sup>	<a href="https://www.aclweb.org/anthology/P19-1544.pdf">https://www.aclweb.org/anthology/P19-1544.pdf</a>	<a href="https://github.com/ZephyrChenzf/SF-ID-Network-For-NLU">https://github.com/ZephyrChenzf/SF-ID-Network-For-NLU</a>
13	CAPSULE-NLU <sup>[39]</sup>	<a href="https://www.aclweb.org/anthology/P19-1519.pdf">https://www.aclweb.org/anthology/P19-1519.pdf</a>	<a href="https://github.com/czhang99/Capsule-NLU">https://github.com/czhang99/Capsule-NLU</a>

## 7 未来的研究方向与展望

随着深度学习算法的不断发展, 以深度学习为主要手段的人工智能技术在自然语言处理领域得到了广泛的应用. 人机对话也不再只局限于简单的控制命令了, 而是能够进一步地实现人类本身的自然语言的理解, 这主要归功于全世界的研究者在深度学习上做出的巨大贡献. 口语理解是人机对话中的关键任务, 从早期的 SVM<sup>[21]</sup>, CRF<sup>[28]</sup>, RecNN<sup>[30]</sup>, RNN<sup>[25]</sup>等独立建模方法, 到目前主流的 Self-attentive Gate Model<sup>[13]</sup>, SF-ID network<sup>[15]</sup>, Stack-propagation<sup>[37]</sup>, BERT SLU<sup>[61]</sup>等联合建模方法, 模型的表示学习能力和任务性能都得到了很大的提升. 但是, 许多关键性问题仍值得深入探索. 本文总结了以下 5 个有待研究的方向.

- (1) 意图与槽位任务交互的错误传播问题. 目前, 意图识别与槽位填充有联合建模方法效果较好. 它们采用了门控机制、直接隐向量或迭代机制进行任务之间的交互. 但是在交互过程中, 一旦句子的意图识别错误, 就会引入错误的意图信息, 误导所有单词的槽位预测; 反之也一样. 这种错误传播问题使得模型训练变得困难, 并且无法快速收敛. 因此, 未来意图与槽位任务交互的错误传播问题, 仍会是学者们需要研究和攻克的一大难题.
- (2) 模型的容错性问题. 口语文本是由自动语音识别技术(automatic speech recognition, ASR)生成的, 常常伴随着一些错别字, 这些错别字可能会对意图识别有影响; 另外, 如果这些错别字是实体的一部分, 槽位预测就不可能正确. 目前所提出的这些模型都忽略了一个问题, 认为这些文本是无别字的. 因此, 如何采用有效的方式去解决文本中出现大量错别字时模型的容错性问题, 是 SLU 任务发展过程中需要研究的重要问题.
- (3) 通过强化学习构建模型表示方法. 强化学习<sup>[87]</sup>与环境进行交互获取奖赏值来改变行为的学习方式, 更像是人类学习时的经验总结. 在自然语言处理领域, 强化学习在文本分类<sup>[88]</sup>、关系抽取<sup>[89]</sup>、对话生成等方面都取得了不错的成绩. 口语理解任务也可以通过强化学习方法构建策略模型和价值模型学习句子表示. 目前, 强化学习在口语理解任务上还未进行探索.

- (4) 构建多意图识别和槽位填充的联合模型. 口语理解中, 有时候一句话可能包含多个意图, 比如帮我预定某某酒店, 现在打个滴滴快车过去. 这句话包含两个意图: 预定酒店和打车. 我们在日常生活中这种表达方式还是比较常见的, 而目前的方法只能分类单个意图. 因此, 如何构建多意图识别和槽位填充的联合模型, 也是未来亟待解决的问题之一.
- (5) 基于预训练最新进展的联合建模方法. 近年来, 预训练模型得到了广泛的发展, 不仅在学术界得到了认可, 在工业界也被应用到具体的产品上, 大大提升了用户体验. 它是在大规模语料上训练得到的, 迁移到其他任务上性能表现优异; 并且采用了 WordPiece/BPE 的词分割算法, 可以有效地解决 OOV 问题. 整个模型是基于 Transformer 架构的, 能够实现并行计算. 本文中, Chen 等人<sup>[61]</sup>和 Qin 等人<sup>[37]</sup>介绍了基于早期提出的预训练模型 BERT 构建的口语理解的联合建模方法, 对于之后提出的自回归模型(XLNet)和自编码模型(RoBERTa, ELECTRA, ALBERT 等)<sup>[90]</sup>, 还未进行探索.

#### References:

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436–444.
- [2] Chen C, Zhu QQ, Yan R, Liu JF. Survey on deep learning based open domain dialogue system. *Chinese Journal of Computers*, 2019, 42(7): 1439–1466 (in Chinese with English abstract).
- [3] López-Cózar R, Callejas Z, Griol D, *et al.* Review of spoken dialogue systems. *Loquens*, 2014, 1(2): 2386–2637.
- [4] 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).
- [5] Chen H, Liu X, Yin D, *et al.* A survey on dialogue systems: Recent advances and new frontiers. *ACM SIGKDD Explorations Newsletter*, 2017, 19(2): 25–35.
- [6] Gao J, Galley M, Li L. Neural approaches to conversational AI. *Foundations and Trends® in Information Retrieval*, 2019, 13(2–3): 127–298.
- [7] Zhang X, Wang H. 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*. 2016. 2993–2999.
- [8] Guo D, Tur G, Yih W, *et al.* Joint semantic utterance classification and slot filling with recursive neural networks. In: *Proc. of the IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2014. 554–559.
- [9] Liu B, Lane I. Attention-based recurrent neural network models for joint intent detection and slot filling. arXiv:1609.01454, 2016.
- [10] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate ICLR. In: *Proc. of the Int'l Conf. on Learning Representations*. 2015.
- [11] Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. In: *Proc. of the Conf. on Empirical Methods in Natural Language Processing*. 2015. 1412–1421.
- [12] Goo CW, Gao G, Hsu YK, *et al.* Slot-gated modeling for joint slot filling and intent prediction. In: *Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.2*. 2018. 753–757.
- [13] Li C, Li L, Qi J. A self-attentive model with gate mechanism for spoken language understanding. In: *Proc. of the Conf. on Empirical Methods in Natural Language Processing*. 2018. 3824–3833.
- [14] Chen M, Zeng J, Lou J. A self-attention joint model for spoken language understanding in situational dialog applications. arXiv:1905.11393, 2019.
- [15] E HH, Niu P, Chen Z, *et al.* 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*. 2019. 5467–5471.
- [16] Gorin AL, Riccardi G, Wright JH. How may I help you? *Speech Communication*, 1997, 23(1-2): 113–127.
- [17] Price P. Evaluation of spoken language systems: The ATIS domain. In: *Proc. of the Speech and Natural Language: Proc. of a Workshop Held at Hidden Valley*. 1990. 91–95.
- [18] De Mori R. Spoken language understanding: A survey. In: *Proc. of the IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*. IEEE, 2007. 365–376.
- [19] Young S. Talking to machines (statistically speaking). In: *Proc. of the 7th Int'l Conf. on Spoken Language Processing*. 2002.

- [20] Hahn S, Dinarelli M, Raymond C, *et al.* Comparing stochastic approaches to spoken language understanding in multiple languages. *IEEE Trans. on Audio, Speech, and Language Processing*, 2010, 19(6): 1569–1583.
- [21] Haffner P, Tur G, Wright JH. Optimizing SVMs for complex call classification. In: *Proc. of the IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, Vol.1. IEEE, 2003. 632–635.
- [22] Schapire RE, Singer Y. BoosTexter: A boosting-based system for text categorization. *Machine Learning*, 2000, 39(2-3): 135–168.
- [23] Fu B, Liu T. Implicit user consumption intent recognition in social media. *Ruan Jian Xue Bao/Journal of Software*, 2016, 27(11): 2843–2854 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4870.htm> [doi: 10.13328/j.cnki.jos.004870]
- [24] Deoras A, Sarikaya R. Deep belief network based semantic taggers for spoken language understanding. In: *Proc. of the Interspeech*. 2013. 2713–2717.
- [25] Ravuri S, Stolcke A. Recurrent neural network and LSTM models for lexical utterance classification. In: *Proc. of the 16th Annual Conf. of the Int'l Speech Communication Association*. 2015. 135–139.
- [26] Zhang ZC, Zhang ZW, Zhang ZM. User intent classification based on IndRNN-attention. *Journal of Computer Research and Development*, 2019, 56(7): 1517–1524 (in Chinese with English abstract).
- [27] Zhou JZ, Zhu ZK, He ZQ, Chen WL, Zhang M. Hybrid neural network models for human-machine dialogue intention classification. *Ruan Jian Xue Bao/Journal of Software*, 2019, 30(11): 3313–3325 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5862.htm> [doi: 10.13328/j.cnki.jos.005862]
- [28] Raymond C, Riccardi G. Generative and discriminative algorithms for spoken language understanding. In: *Proc. of the 8th Annual Conf. of the Int'l Speech Communication Association*. 2007. 1605–1608.
- [29] McCallum A, Freitag D, Pereira FCN. Maximum entropy Markov models for information extraction and segmentation. In: *Proc. of the 17th Int'l Conf. on Machine Learning*. 2000. 591–598.
- [30] Yao K, Zweig G, Hwang MY, *et al.* Recurrent neural networks for language understanding. In: *Proc. of the Interspeech*. 2013. 2524–2528.
- [31] Yao K, Peng B, Zweig G, *et al.* Recurrent conditional random field for language understanding. In: *Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014. 4077–4081.
- [32] Mesnil G, He X, Deng L, *et al.* Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding. In: *Proc. of the Interspeech*. 2013. 3771–3775.
- [33] Yao K, Peng B, Zhang Y, *et al.* Spoken language understanding using long short-term memory neural networks. In: *Proc. of the IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2014. 189–194.
- [34] Simonnet E, Camelin N, Deléglise P, *et al.* Exploring the use of attention-based recurrent neural networks for spoken language understanding. In: *Proc. of the Machine Learning for Spoken Language Understanding and Interaction NIPS 2015 Workshop (SLUNIPS 2015)*. Montreal, 2015. 642–650.
- [35] Zhu S, Yu K. Encoder-decoder with focus-mechanism for sequence labelling based spoken language understanding. In: *Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017. 5675–5679.
- [36] Hakkani-Tür D, Tür G, Celikyilmaz A, *et al.* Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM. In: *Proc. of the Interspeech*. 2016. 715–719.
- [37] Qin L, Che W, Li Y, *et al.* A stack-propagation framework with token-level intent detection for spoken language understanding. In: *Proc. of the Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP)*. 2019. 2078–2087.
- [38] Wang Y, Shen Y, Jin H. A Bi-model based RNN semantic frame parsing model for intent detection and slot filling. In: *Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol.2. 2018. 309–314.
- [39] Zhang C, Li Y, Du N, *et al.* Joint slot filling and intent detection via capsule neural networks. In: *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019. 5259–5267.
- [40] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2014. 580–587.
- [41] Girshick R. Fast R-CNN. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2015. 1440–1448.

- [42] Ren S, He K, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. In: Proc. of the Advances in Neural Information Processing Systems. 2015. 91–99.
- [43] Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 2980–2988.
- [44] Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. In: Proc. of the European Conf. on Computer Vision. Cham: Springer, 2016. 21–37.
- [45] Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv:1804.02767, 2018.
- [46] Li YG, Zhou XG, Sun Y, Zhang HG. Research and implementation of Chinese microblog sentiment classification. Ruan Jian Xue Bao/Journal of Software, 2017, 28(12): 3183–3205 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5283.htm> [doi: 10.13328/j.cnki.jos.005283]
- [47] Wu RS, Wang HL, Wang ZQ, Zhou GD. Short text summary generation with global self-matching mechanism. Ruan Jian Xue Bao/Journal of Software, 2019, 30(9): 2705–2717 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5850.htm> [doi: 10.13328/j.cnki.jos.005850]
- [48] Qiu M, Li FL, Wang S, *et al.* Alime chat: A sequence to sequence and rerank based chatbot engine. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol.2. 2017. 498–503.
- [49] Xu H, Zhang H, Han K, *et al.* Learning alignment for multimodal emotion recognition from speech. arXiv:1909.05645, 2019.
- [50] Codevilla F, Miiller M, López A, *et al.* End-to-end driving via conditional imitation learning. In: Proc. of the IEEE Int'l Conf. on Robotics and Automation (ICRA). IEEE, 2018. 1–9.
- [51] Badue C, Guidolini R, Carneiro RV, *et al.* Self-driving cars: A survey. arXiv:1901.04407, 2019.
- [52] Leydesdorff L, Vaughan L. Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. Journal of the American Society for Information Science and, Technology, 2006, 57(12): 1616–1628.
- [53] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research, 2003, 3: 993–1022.
- [54] Deerwester S, Dumais ST, Furnas GW, *et al.* Indexing by latent semantic analysis. Journal of the American Society for Information Science, 1990, 41(6): 391–407.
- [55] Bengio Y, Ducharme R, Vincent P, *et al.* A neural probabilistic language model. Journal of Machine Learning Research, 2003, 3: 1137–1155.
- [56] Mikolov T, Kombrink S, Deoras A, *et al.* RNNLM-recurrent neural network language modeling toolkit. In: Proc. of the ASRU Workshop. 2011. 196–201.
- [57] Mikolov T, Chen K, Corrado G, *et al.* Efficient estimation of word representations in vector space. arXiv:1301.3781, 2013.
- [58] Mikolov T, Sutskever I, Chen K, *et al.* Distributed representations of words and phrases and their compositionality. In: Proc. of the Advances in Neural Information Processing Systems. 2013. 3111–3119.
- [59] Mikolov T. word2vec. 2013. <https://code.google.com/archive/p/word2vec.2013>.
- [60] Xu P, Sarikaya R. Convolutional neural network based triangular CRF for joint intent detection and slot filling. In: Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding. IEEE, 2013. 78–83.
- [61] Chen Q, Zhuo Z, Wang W. Bert for joint intent classification and slot filling. arXiv:1902.10909, 2019.
- [62] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back propagating errors. Nature, 1986, 323: 533–536.
- [63] Mikolov T, Karafiát M, Burget L, *et al.* Recurrent neural network based language model. In: Proc. of the 11th Annual Conf. of the Int'l Speech Communication Association. 2010. 1045–1048.
- [64] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780.
- [65] Cho K, van Merriënboer B, Bahdanau D, *et al.* On the properties of neural machine translation: Encoder–decoder approaches. In: Proc. of the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8). 2014. 103–111.
- [66] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. In: Proc. of the Advances in Neural Information Processing Systems. 2017. 5998–6008.
- [67] Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. In: Proc. of the Advances in Neural Information Processing Systems. 2017. 3856–3866.

- [68] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proc. of the Advances in Neural Information Processing Systems. 2012. 1097–1105.
- [69] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 1–9.
- [70] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 770–778.
- [71] Hemphill CT, Godfrey JJ, Doddington GR. The ATIS spoken language systems pilot corpus. In: Proc. of the Speech and Natural Language: Proc. of a Workshop Held at Hidden Valley. Pennsylvania, 1990. 96–101.
- [72] Coucke A, Saade A, Ball A, *et al.* Snips voice platform: An embedded spoken language understanding system for private-by-design voice interfaces. arXiv:1805.10190, 2018.
- [73] Lai S, Xu L, Liu K, *et al.* Recurrent convolutional neural networks for text classification. In: Proc. of the 29th AAAI Conf. on Artificial Intelligence. 2015. 2267–2273.
- [74] Boden M. A guide to recurrent neural networks and backpropagation. The Dallas Project, 2002, 2: 1–10.
- [75] Luong MT, Le QV, Sutskever I, *et al.* Multi-task sequence to sequence learning. arXiv:1511.06114, 2015.
- [76] Venugopalan S, Rohrbach M, Donahue J, *et al.* Sequence to sequence-video to text. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 4534–4542.
- [77] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 2005, 18(5-6): 602–610.
- [78] Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Proc. of the Neural Information Processing Systems. Montreal, 2014. 3104–3112.
- [79] Graves A, Jaitly N, Mohamed A. Hybrid speech recognition with deep bidirectional LSTM. In: Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding. IEEE, 2013. 273–278.
- [80] Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1. 2019. 4171–4186.
- [81] Wu Y, Schuster M, Chen Z, *et al.* Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv:1609.08144, 2016.
- [82] Forney GD. The viterbi algorithm. Proc. of the IEEE, 1973, 61(3): 268–278.
- [83] Kim Y, Jernite Y, Sontag D, *et al.* Character-aware neural language models. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. 2016. 2741–2749.
- [84] Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv:1510.03820, 2015.
- [85] Jeong M, Lee GG. Triangular-chain conditional random fields. IEEE Trans. on Audio, Speech, and Language Processing, 2008, 16(7): 1287–1302.
- [86] E HH, Zhang WJ, Xiao SQ, Cheng R, Hu YX, Zhou XS, Niu PQ. Survey of entity relationship extraction based on deep learning. Ruan Jian Xue Bao/Journal of Software, 2019, 30(6): 1793–1818 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5817.htm> [doi: 10.13328/j.cnki.jos.005817]
- [87] Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT Press, 2018.
- [88] Zhang T, Huang M, Zhao L. Learning structured representation for text classification via reinforcement learning. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. 2018. 6053–6060.
- [89] Zeng X, He S, Liu K, *et al.* Large scaled relation extraction with reinforcement learning. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. 2018. 5658–5665.
- [90] Qiu X, Sun T, Xu Y, *et al.* Pre-trained models for natural language processing: A survey. Science China Technological Sciences, 2020, 63: 1872–1897.



## 附中文参考文献:

- [2] 陈晨, 朱晴晴, 严睿, 柳军飞. 基于深度学习的开放领域对话系统研究综述. 计算机学报, 2019, 42(7): 1439–1466.
- [4] 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.
- [23] 付博, 刘挺. 社交媒体中用户的隐式消费意图识别. 软件学报, 2016, 27(11): 2843–2854. <http://www.jos.org.cn/1000-9825/4870.htm> [doi: 10.13328/j.cnki.jos.004870]
- [26] 张志昌, 张珍文, 张治满. 基于 IndRNN-Attention 的用户意图分类. 计算机研究与发展, 2019, 56(7): 1517–1524.
- [27] 周俊佐, 朱宗奎, 何正球, 陈文亮, 张民. 面向人机对话意图分类的混合神经网络模型. 软件学报, 2019, 30(11): 3313–3325. <http://www.jos.org.cn/1000-9825/5862.htm> [doi: 10.13328/j.cnki.jos.005862]
- [46] 李勇敢, 周学广, 孙艳, 张焕国. 中文微博情感分析研究与实现. 软件学报, 2017, 28(12): 3183–3205. <http://www.jos.org.cn/1000-9825/5283.htm> [doi: 10.13328/j.cnki.jos.005283]
- [47] 吴仁守, 王红玲, 王中卿, 周国栋. 全局自匹配机制的短文本摘要生成方法. 软件学报, 2019, 30(9): 2705–2717. <http://www.jos.org.cn/1000-9825/5850.htm> [doi: 10.13328/j.cnki.jos.005850]
- [86] 鄂海红, 张文静, 肖思琪, 程瑞, 胡莺夕, 周筱松, 牛佩晴. 深度学习实体关系抽取研究综述. 软件学报, 2019, 30(6): 1793–1818. <http://www.jos.org.cn/1000-9825/5817.htm> [doi: 10.13328/j.cnki.jos.005817]



魏鹏飞(1991—), 男, 助理实验师, 主要研究领域为自然语言理解, 任务对话系统, 强化学习.



汪明慧(1974—), 女, 博士, 讲师, 主要研究领域为人工智能.



曾碧(1963—), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为人工智能, 智能人机交互, 智能机器人.



曾安(1978—), 女, 博士, 教授, CCF 高级会员, 主要研究领域为人工智能, 机器学习.