

问句意图理解与识别技术研究进展

韦婷婷 葛晓月 宋世领 熊俊涛*

(华南农业大学数学与信息学院 广东 广州 510642)

摘要 问句意图理解与识别是问答和对话系统的重要组成部分,包含上下文语义理解及意图分类子任务。尽管问句意图理解和识别任务已取得大量研究成果,但是目前仍缺少相关系统性的回顾与总结的研究工作。因此,梳理近些年的研究,从问句意图理解与识别的概念出发,介绍其研究目的与意义,总结归纳评测方法与数据来源,刻画方法间的代际差异,并重点阐述目前流行的深度学习方法在问句意图理解与识别上的研究现状,探讨其在未来潜在的发展方向。

关键词 问答系统 问句意图理解与识别 文本分类 深度学习

中图分类号 TP391 文献标志码 A DOI:10.3969/j.issn.1000-386x.2025.09.002

RESEARCH PROGRESS OF QUESTION INTENT UNDERSTANDING AND DETERMINATION TECHNOLOGY

Wei Tingting Ge Xiaoyue Song Shiling Xiong Juntao*

(College of Mathematical Sciences and Information, South China Agricultural University, Guangzhou 510642, Guangdong, China)

Abstract Question-intent understanding and recognition is an important part of question-answer and dialogue systems, which includes contextual semantic understanding and intention classification subtasks. Although a large number of research results have been achieved in the task of question intention determination, there is still a lack of systematic review and summary of related research work. Therefore, this paper compares the recent research, starts from the concept of question intent understanding and determination, introduces its research purpose and significance, summarizes the evaluation methods and data sources, portrays the generational differences between methods, and focuses on the application of currently popular deep learning methods to question intent understanding and determination, and finally explores its potential development direction in the future.

Keywords Q&A systems Question intent understanding and determination Text classification Deep learning

0 引言

问句意图理解是问答和对话系统的先导任务^[1],也是后续意图识别、对话管理和自然语言生成等任务的基础。随着语言情境的复杂多变,如何快速高效地处理、理解、分类和优化用户的意图,使得语义理解更加精准,是问答和对话系统进入全面化应用的关键。

但是在实际应用中,问句意图理解和识别的难点在于如何让计算机理解用户所输入的自然语言表述。例如,用户输入“苹果会卡顿吗?”,系统需要理解这句话的上下文语义信息,准确识别出用户所问的“苹果”是属于数码品牌而非水果的类别,才能快速并精确地返回令用户满意的答案。问句意图理解与识别使问答和对话系统能够更好地进行答案检索工作^[2],因而对性能起着关键性作用。图1为问句意图理解与识别在问

答系统中的应用实例^[3]。

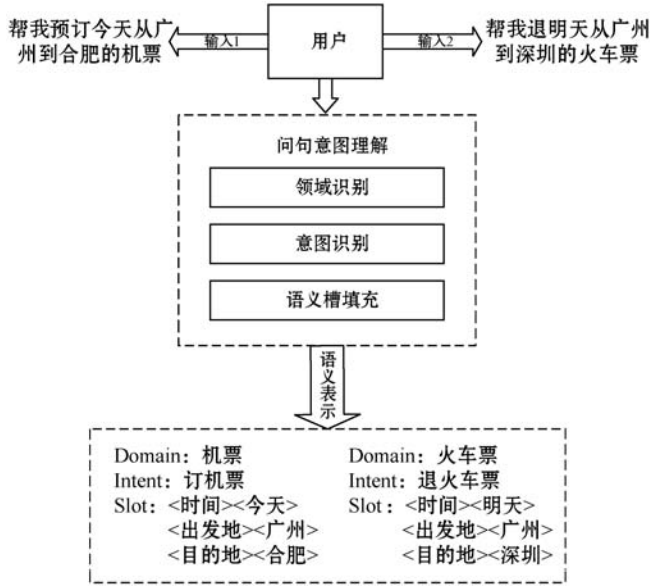


图1 问答系统基本构造实例^[3]

在大规模应用深度学习之前,问句意图理解与识别以传统的机器学习方法为主,准确率一直在缓慢提升。深度学习的引入和发展使问句意图理解与识别任务的准确率得到了很大的提高^[4]。随着后续注意力、LSTM模型和BERT等模型的引入与研究,问句意图理解与识别方法目前在英文数据集上的F1值已达到约96%。中文数据集方面,在2018年全国知识图谱与语义计算大会(CCKS2018)的技术评测中,其F1值也达到了88.35%以上^[5]。

尽管问句意图理解与识别已取得大量研究成果,但是目前仍缺少对有关问句意图理解与识别方法进行系统性的回顾与总结的研究工作。因此,本文将系统地刻画问句意图理解与识别方法的代际差异,通过对比和分析不同方法下研究结果的不同,探讨其在未来潜在发展方向,为科研人员厘清问句意图理解与识别研究发展现状具有重要意义。

1 定义及评测方法

1.1 问句意图理解与识别的定义

传统搜索引擎中,Broder^[6]将用户查询意图分为导航型、信息型和事务型三类。导航型是指用户获取某个网页URL的意图,比如查询学校官网等;信息型是指用户查询数据库中存在的信息的意图,除了阅读信息没有进一步互动,比如在百度百科查询某些资料;事务型是指用户通过系统交互完成一项任务的意图,比如在线购物、下载歌曲等。

与传统搜索不同,目前的问答系统追求更为精准

的答案。因此,对于用户的查询意图需要划分为更加具体的特定类别。问答系统中,问句意图理解主要任务是将用户输入的自然语言问句转化为机器可以理解的结构化表示。问句意图识别指的是为用户所提问题赋予一个预先定义类别标签(如询问的是“航班”还是“航线”,两者属于不同的意图类型),从而方便系统寻找答案所属范畴。问句意图理解与识别方法遵循的一般流程是首先对问句进行数据标注和预处理,再将数据集进行特征提取与向量化,而后利用分类模型(例如回归模型、支持向量机、神经网络等)进行数据训练,构造分类器。

1.2 问句意图理解与识别的评测方法

一个合适的评测方法,能够更好地衡量模型的准确率。本文根据相关理论与文献,整理出问句意图理解与识别任务的主要评分指标。问句意图理解与识别的评测方法主要有精准率(Accuracy)、准确率(Precision)、召回率(Recall)和F1测度值(F1-Measure)等^[7]。根据分类结果可建立混合矩阵如表1所示。

表1 分类结果混合矩阵

	实际属于该意图类别的样本数	实际不属于该意图类别的样本数
系统判断属于该意图类别	a	b
系统判断不属于该意图类别	c	d

(1) 精准率。精准率是分类正确的正例数据占分类为正例数据的比例。计算式如下:

$$A = \frac{\text{正确分类的样本数}}{\text{分类样本总数}} \times 100\% = \frac{a + d}{a + b + c + d} \times 100\% \quad (1)$$

精度测度指标比较易于理解,但存在一定局限性。精准率在评价处理极度倾斜(skewed)的数据的任务时效果较差。

(2) 准确率。准确率是最常见的评价指标。准确率是系统正确判断的样本数与系统判断的样本总数的比率,体现了系统识别结果的准确程度。计算式如下:

$$P = \frac{\text{正确分类的样本数}}{\text{属于该类的样本总数}} \times 100\% = \frac{a}{a + b} \times 100\% \quad (2)$$

(3) 召回率。召回率是系统正确判断为该意图类别的样本数与应属于该类的样本总数的比率,体现了意图识别结果的完备性。计算式如下:

$$R = \frac{\text{正确分类的样本数}}{\text{实际该类别的样本总数}} \times 100\% = \frac{a}{a + c} \times 100\% \quad (3)$$

(4) F1值。准确率和召回率有时候会出现矛盾的情况,因此需要综合考虑,最常见的方法就是使用综

合评价指标分数。综合评价指标是准确率和召回率的调和平均值,其数学式为:

$$F = \frac{(a^2 + 1)P \times R}{a^2(P + R)} \quad (4)$$

当参数 $a = 1$ 时,就是最常见的 F1 测度值:

$$F_1 = \frac{2PR}{P + R} \quad (5)$$

一般情况下, F1 越高则说明该方法就越有效。

2 数据来源

问句意图理解与识别任务使用的数据集来源较为广泛,可以使用互联网上公开开源数据集,比如微软发布的源自 Bing 搜索引擎中用户查询的数据集 MS MARCO 以及源自维基百科文章的问答对数据集 SQuAD;也可以通过爬虫等技术从相关网站上收集所需要的公开数据。表 2 梳理了近年来相关研究中使用的中英文数据集名称、类别、样本数量、代表文献。

表 2 问句意图理解与识别常用数据集及来源

数据集	数据来源	样本数量
ATIS (英文)	国防高级研究计划局 http://lisaweb.iro.umontreal.ca/transfert/lisa/users/mesnilgr/atis	5 871
Snips (英文)	Snips 个人语音助理 https://console.snips.ai	14 484
SMP (中文)	全国社交媒体处理大会 SMP2017 中文人机对话技术评测 (ECDT) 任务一 http://ir.hit.edu.cn/smp2017ecdt-data	966
CCKS (中文)	全国知识图谱与语义计算大会技术评测任务 https://biendata.com/CCKS2021	12 000
百度 BroadAI (中英文)	百度 ai.baidu.com/broad	200 000
旅行问答数据集 (中文)	马蜂窝、携程问答板块	48 012
医药咨询数据集 (中文)	39 问答网、有问必答网	225 640

2.1 ATIS 数据集 (英文)

20 世纪 90 年代初,国防高级研究计划局公布的航空旅游信息系统的日志数据 (Airline Travel Information System, ATIS) 是意图识别任务中使用最多的经典数据集之一,主要内容是客户询问航班有关的录音集,该数据集含有 4 978 条训练集、893 条测试集,由以 In/Out/Begin (IOB) 格式表示的句子组成^[5],同时标注有对应槽位、word2id、槽位 id。

表 3 列出了 ATIS 数据集中几种不同类别的数据样式示例。

表 3 ATIS 数据集示例

语料	意图类别	属性值	实体
Find flight arriving new york city next saturday	航班	目的地:new york city	toloc. city_name
what airline is aa	航线	航线编号:aa	airline-code

2.2 百度 BroadAI 数据集 (中英文)

2017 年百度推出目前全球规模最大的“BROAD”百度 AI 公开数据集 (Baidu Research Open-Access Dataset),适用任务包括室外场景理解、视频精彩片段、阅读理解三类。以百度阅读理解 DuReader 数据集为例,该数据集收集了 20 多万条问题以及 42 万条人工生成的答案。DuReader 数据集标记了问题类型。表 4 列出了 DuReader 数据集中几种不同类别的数据样式示例。

表 4 DuReader 数据集示例

问句类型	问句
Yes-No	上海迪士尼可以带吃的进去吗 Can Shanghai Disney bring food in
Description	丰田卡罗拉怎么样 How is Toyota Carola
Entity	iPhone 在哪天发布 On which day will iPhone be released

2.3 CCKS 数据集 (中文)

全国知识图谱与语义计算大会 (CCKS: China Conference on Knowledge Graph and Semantic Computing) 由中国中文信息学会语言与知识计算专业委员会主办,其会议数据集常用于知识表示、自然语言理解、智能问答等相关技术领域研究。以 2021 年 CCKS 中的一个电信领域数据集为例,该数据集由阿里巴巴与中国移动在线服务联合提供,包含 5 000 条训练数据、1 000 条测试数据。表 5 列出了 CCKS2021 电信数据集中的几种数据样式示例。

表 5 CCKS2021 电信数据集示例

问句	意图	实体	约束属性名	约束属性值
9 元百度专属定向流量包如何取消	取消方式	专属定向流量包	价格 子业务	9 百度
220 元移动电视盒子应如何办理	开通方式	互联网电视	价格	220
什么时候可以办理 3 元流量日包	上线时间	流量日包	价格	3

2.4 Snips 语音数据集 (英文)

该数据集是通过 Snips 个人语音助手收集的,包含 13 784 个训练和 700 个测试样本,该训练集有 72 个

槽标签和 7 种意图类型。表 6 列出了 Snips 数据集中的几种数据样式示例。

表 6 Snips 数据集示例

问句	意图类别
What's the weather forecast for my current spot	GetWeather
Play the lamentation of cloris	PlayMusic
I am giving this current book album 0 out of 6 stars	RateBook

2.5 SMP 数据集(中文)

SMP 中文人机对话技术评测由哈工大、清华大学和科大讯飞自 2017 年开始联合承办,该中文数据集涉及股票、点歌、数学、邮件和天气等多个问题。以 SMP2018 用户意图分类(任务一)数据集为例,训练集有 2 299 条样本,测试集有 667 条样本,已开源供研究使用。表 7 列出了数据样式示例。

表 7 SMP 数据集示例

问句	意图类别
中国银行股票怎么样	股票
帮我写一篇邮件	邮件
你最近在干嘛呢	闲聊

3 问句意图理解与识别发展历程

根据研究方法的变迁,问句意图理解与识别从最初的基于规则模板,到基于统计学,而后是基于机器学习的方法,到如今以深度学习方法为主流。本文通过检索近 15 年内的中、英文文献,分别绘制出如下中文(图 2)、英文(图 3)文献统计图,以直观看不同年份、不同阶段下问句意图理解与识别相关的发文数量,并标注了三个发展阶段:阶段一为基于规则模板的方法,阶段二为基于统计学或机器学习的方法,阶段三为基于深度学习的方法。可以看出,目前该领域越来越受众多学者的关注,且正在持续发展中。以下将分别介绍不同阶段的主要方法。

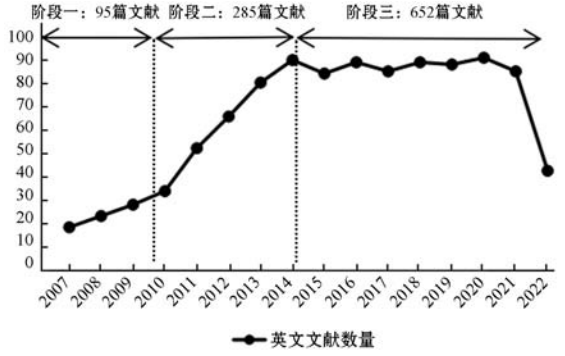


图 3 英文文献统计图

3.1 基于规则的方法

传统的问句意图理解与识别方法是将意图识别视为语义话语分类的问题进行研究,多用于单轮对话系统中意图的检测。最早的问句意图理解与识别使用基于规则模板的方法实现,主要包含词表穷举法和规则模板解析两类。

词表穷举法主要通过词表直接匹配的方式来获取查询意图,同时也可以加入比较简单、查询模式较集中的类别。其方式实现较为简单,能够较准确地解决高频词。但该方法过于依赖人工参与,因此较为死板,很难实现自动化,更无法处理较多的意图和长尾问题,同时在使用时文本必须全部命中穷举集才有效果,因此模型的召回率较低,精确率极高。词表穷举法在问句意图理解与识别任务上很难大规模应用。

规则模板解析方法是指将用户问句与模板库中的句型模板进行匹配,然后根据句型模板把问句的语义信息提取出来。该方法实现需要分词、词性标注、命名实体识别、依存句法分析、语义分析等步骤,为后来逐渐先进的意图识别方法提供了雏形。唐素勤等^[8]引入概念关系知识表示模型,先使用模糊集的方法通过问句意图理解模块提取问句涉及的概念、属性及关系,然后通过意图识别模块模板的模糊匹配实现结果的反馈。模板匹配方法是基于字符表层的文本分析技术,因此具有只能处理各个关键词中有分隔词的问句的局限性。曹庆花^[9]引入了本体技术构建知识库,在问句意图理解模块使用句法分析处理将自然语言问句表达成树状结构数据,然后与 FAQ 问答数据库中预先储存的“问句-答案”对进行匹配实现问句意图识别。规则模板解析方法适合应用于较基础、规则性较强的智能意图识别情境,如搜狗通用搜索、智能物联网等。

规则模板解析法在数据上使用字典结构,因此查询速度较快,对规则性的问题有较好的意图识别效果,能够较为准确地提取信息。在基于规则模板解析方法中,句型模板的数量与规模在一定程度上影响着问句意图理解与识别的效率,因此当文本信息与

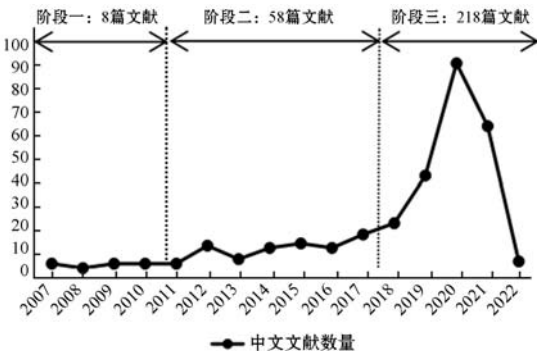


图 2 中文文献统计图

模板数量不断增加,基于规则模板的方法不仅耗费大量的人力、物力与时间成本,而且还有规则覆盖率小、多领域适用性差、无法实现信息的动态更新等缺陷。

3.2 基于统计学与机器学习的方法

问句意图理解与识别问题具有类别多、特征之间有较大关联性等特点。面对日益增长的海量的、多样化的文本数据,早期基于规则的分类方法逐渐到达了瓶颈。随着人工智能技术的高速发展,这使得许多学者开始探索将机器学习技术迁移应用到意图识别领域中,基于统计学与机器学习的方法百花齐放。传统的机器学习模型通常是利用统计学原理,从预先标记好的语料中学习其内在的分类规律,然后进行类别检测。

前期阶段主要使用传统统计学与机器学习方法。Kim等^[10]提出一种基于降维思路的多意图识别方法。实验使用了支持向量机(SVM)、kNN和重心向量分类(centroid-based classification)三种方法测试了降低维度后的意图识别效率。实验证明该方法有效解决了文本输入空间的高维性、文档向量的稀疏性、容易丢失全局特征的多意图识别问题,并使用SVM方法在Reuter-21578数据集上取得了88.9%的准确率。赵小华^[11]改进了TFIDF算法,用特征选择函数代替传统公式中的反文档频率IDF,结合K近邻分类方法(KNN)识别问句意图。在问句意图理解模块,从文本中抽取可以代表句子特征的字、词或短语并使用n-grams特征表示法,以n个字为单位表示特征值。该方法有效解决了关联特征词的提取问题,但特征数较多或样本分布不平衡时分类效果较差。然而,传统的统计学或机器学习方法在处理文本时会有处理速度慢、耗时较长和处理缺失数据困难等缺点,极大影响了分类效率。

后期阶段,众多学者致力于改善传统方法与模型,结合各种统计学公式模型与机器学习方法形成流水线网络进一步提高意图识别的模型性能。董云耀等^[12]将支持向量机与粗糙集理论(RST)结合,提出一种中文意图分类方法——CRV进行问句意图识别,实验选择主要的疑问词、谓语句、中心名词及其之间的依存关系作为问句的特征项,在问句意图理解模块将问句表示成计算机可以理解的问句向量形式。实验表明该方法在提高分类精度的同时也降低了数据空间维度。在自行收集的计算机网络网课中学生问题数据集上实验,与SVM模型相比,数据空间维数从1200降至923,准确率提高0.67%。康海燕等^[13]提出了一种改进的贝叶斯和疑问词相结合的方法,在用户初次查询时,使用疑问词提取、关键词提取和扩展、问句类型提取和新词识别等一系列问句表征方法对问句信息进行

形式化表示,从而充分理解问句意图,实现意图的动态更新,但该方法对输入数据的表达形式较敏感。

Li等^[14]提出了一种基于动态学习机制和相似性匹配算法的问句意图理解与识别方法,使用ICTCLAS软件工具包进行词性标注和未知词识别,系统通过问句相关要素与预先收集好的语句要素库匹配,要素超过某一阈值的被判定为复杂问题,然后通过对复杂问句的进一步处理,如通过使用问答系统修改问句或提供相关关键词指导等,可以大大提高复杂问句的问答效率,有效解决了用户问句中复杂意图的理解与识别问题,提高了搜索引擎的查询质量。

洪宗祥等^[15]在问句意图理解模块将问句用语义网络表示,然后调用查询算法在关系树模型上查询,最后通过模板匹配返回意图识别结果,这种方法有助于隐含意图的识别,但只适用于数据集较小的任务。

Heck等^[16]提出一种基于CRF的无监督最大后验(MAP)自适应方法的关系模型以及一种具有诱导实体提取语法的问句意图理解方法,使用迭代爬行算法生成语句注释并用于CRF语义解析器,最后对来自特定领域的问句样本数据使用最大后验方法来改进解析器,该方法不需要收集数据和人工标注,在维基百科开放语料库上取得了61.02%的准确率。

林博^[17]提出用向量空间表示方法、用户本体模型与隐含狄利克雷分布(LDA)主题模型相结合,以达到基于用户意图的个性化查询、用户模型自动更新的目的,实验表明该方法可以通过意图准确反映出用户的兴趣变化,有效提高了搜索引擎的智能化。在涉及18个领域的搜狗新闻数据集中,准确率约69%。

相比于传统的规则模板方法,基于统计学与机器学习的方法更具泛化能力,模型召回率更高,使得意图识别技术有了较大的发展。但该方法也存在特征表达能力弱、难以解决稀疏矩阵和鲁棒性较差等问题。同时算法依旧依赖大量人工标注好的训练语料,因此具有昂贵的人工成本等缺点。

3.3 基于深度学习的方法

深度学习技术起源于人工神经网络,起初在图像处理领域取得优异的性能表现。后来,学者们发现文本数据和图像数据具有共同的特点:通常都是连续、稠密的,有很强的局部关联性。因此,自然语言处理领域也开始引入深度学习技术,极大推动了很多下游应用的发展。

在问答系统任务中,深度学习相关的研究工作主要集中在问答特征的自动提取层面。如何获取准确表示自然语言问句语义的特征,是问句意图理解与识别

的关键环节。在问句意图理解模块的结构化表示方面,Chen^[18]提出使用卷积神经网络(CNN)在单词向量上进行训练的思路,其中单词向量由静态的预训练向量和动态的应用于特定任务的向量组成,可适用于多个领域的句子级意图识别任务。实验设计了四种不同的生成词向量方法:随机生成的特定维度的词向量、使用预先训练好的谷歌 word2vec 词向量、根据特定任务微调模型参数生成的词向量和以上三种方法结合生成的词向量。实验在 7 个不同数据集上证明,预训练过的词向量对问句意图理解与识别任务的性能提升有很大影响,准确率最高在 MPQA 数据集上达到 89.6%。

Kim^[19]提出的基于 word2vec 的 textCNN 文本分类模型,其中使用 word2vec 方法将问句转化为词向量。该模型可以较好地捕捉特征间的局部相关性,对较短问句文本有着训练快速高效的优点,可以应用于意图理解与识别等自然语言问句的分类,然而,由于受卷积核的大小影响,textCNN 模型可解释性不强,在应用中对于特征重要程度的评估较为吃力,因此很难有针对性地调整具体特征。

在具体特定特征的选取方面,Kalchbrenner 等^[20]在单层 CNN 网络的研究基础上提出动态 CNN(Dynamic Convolutional Neural Network,DCNN)模型,提取句子中经常出现的单词或短句作为句子的特征,模型通过动态池化可以获取句子中距离较远词语间的联系,实验证明 DCNN 模型有效解决了传统 CNN 模型对长文本的处理乏力问题,在推特数据集上取得了 87.4% 的准确率。Hashemi 等^[21]使用 CNN 网络提取问句的向量表示作为意图识别的特征,通过离线训练模型参数,利用已标记的问句意图样本学习参数和分类器,最后用未标记的意图样本在线运行模型得到其意图分类结果,实验证明该方法可以有效简化特征工程的工作,在商业搜索引擎日志数据集上取得了 81.6% 的准确率。

Tang 等^[22]提出了一种使用 GRU 对问句文本进行建模的意图识别模型。该模型首先将文本意图转化为向量,然后利用 CNN 和 LSTM 进行问句格式化表示,模型将整个句子的不同宽度卷积核的向量表示接入一个 Average 层,从而得到句子平均向量表示,使问句意图理解模块可以捕获句子的全局语义表征。最后将平均向量表示输入到 GRU 网络中,得到文档向量表示并输送给 softmax 层,最终计算出标签的概率分布。实验证明 LSTM 网络在建模文本表示方面比多重过滤的 CNN 表现得更好,能够更好地捕捉问句全局特征,充分理解分析用户意图与情感,在 Yelp 2015 数据集上取得了 67.6% 的准确率。

随着预训练语言模型 BERT(Bidirectional Encoder

Representation from Transformers)^[23]在文本语义表示上取得的显著成效,不少学者也开始将其引入到问句意图理解与识别任务中。迟海洋等^[24]构建一种基于 BERT 和 BiGRU-Attention 的混合神经网络模型,在意图理解任务中使用 BERT 预训练得出词向量表示并当作后续模型输入,使用 BiGRU 提取问句特征,再运用 Attention 机制提取关键信息并分配权重,最后使用 Softmax 分类器实现意图识别。实验表明引入 Attention 机制后性能显著提升,在自行收集的国内在线健康医疗网站用户问题数据集上准确率为 89.6%。郑新月等^[25]提出一种基于 BERT 的全连接神经网络(FNN)模型,使用 BERT 模型进行问句的句级别的语义表示,采用 cls 符号(token)对应的向量表示上下文进行问句意图理解,再使用 FNN 进行深层特征提取,学者利用拆解法思想将多分类问题转换成二分类问题,充分利用数据资源的同时,更好地理解用户问句的深层语义特征。实验证明了该模型有效解决了传统模型泛化能力差、深层理解语义乏力和计算成本较高等问题,在自行收集的 16 万条自然问句上准确率达到 94%。

ERNIE(Enhanced Representation through Knowledge Integration)预训练语言模型是在 BERT 模型基础上做的进一步优化,其可以更好地捕捉全局语义信息。郑思露^[5]聚焦于实体信息与上下文的关联性对意图识别的重要性,提出了一种基于 ERNIE 的 BiLSTM 模型,在问句意图理解模块使用 ERNIE 依据不同上下文动态生成词向量,通过 BiLSTM 在上下文捕捉意图,结合胶囊神经网络^[26]编码问句意图在不同维度与真实意图标签间的关系。实验证明该模型有效解决了用户输入的语境匮乏和意图模糊的问题,在 CCKS2018 数据集上的实验准确率达到 88.48%。Saha 等^[27]提出一种 BERT-Cap 模型,使用预训练的 BERT 模型生成词向量表示,然后与胶囊层联合优化来捕捉推特上的语音意图特征,该模型旨在通过评估关键词对特定预测任务的重要性来捕获更细粒度的语义关系,实验证明胶囊神经网络有效提高了 BERT 模型的鲁棒性,在推特公开数据集上取得了 77.52% 的准确率。

另外,还有不少学者将意图识别与语义槽填充这两项原本独立的任务视为可相互补充的任务,因此联合模型也是解决问句意图理解与识别的常用方法。图 4 为问句意图识别与语义槽填充联合处理的一般流程。意图识别任务侧重于预测输入文本的意图,而语义槽填充以提取语义概念作为自然语言的约束为主要目的,即给输入中每个单词分配适当的语义标签。相比于两个任务独立处理,联合模型可共享一个编码器实现两个任务的相互交互,不仅可以扩展原型网络,还

可以进一步提升各自的性能表现。表 8 列出了 5 种代表性意图识别与槽填充的联合方法。

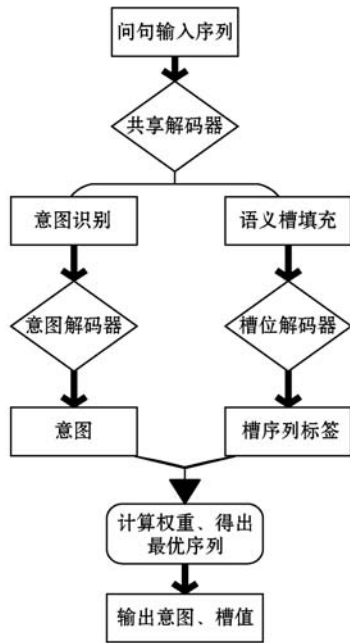


图 4 联合模型一般步骤

表 8 代表性联合模型

方法	基础网络	模型结构	使用方法	数据集	准确率/%	F1 值/%
文献 [28]	双向 RNN、注意力机制	双重异构网络	Adam 优化算法、交叉验证	ATIS	—	95.98
文献 [29]	双向 LSTM、CNN、CRF	双重异构网络	one-hot、softmax 分类器、维特比算法、Adam 优化算法	ATIS	95.9	95.24
文献 [30]	双向 LSTM、注意力机制、槽门机制	三重异构网络	softmax 分类器、Adam 优化算法、显著性检验	ATIS	94.1	95.20
				Snips	91.0	88.80
文献 [31]	BERT	孪生网络	Softmax 分类器、最小化交叉熵、微调模型	ATIS	97.5	—
文献 [32]	胶囊网络	三重同构网络	动态路由、一致性路由算法、最大边际损失函数	ATIS	95.0	95.20
				Snips	97.3	91.80

较早使用联合模型的代表性研究是 Liu 等^[28]提出的基于注意力机制的双向 RNN 模型,实验使用编码器-解码器结构将单词序列映射到槽标签序列,双向 RNN 编码器可以向前和向后读取源单词序列。在 ATIS 数据集上的实验表明,联合模型与两个独立训练模型相比都有一定的提升,证明了联合模型相比于单独训练模型有简化网络、提高性能等显著优势。

进一步地,为了充分利用意图识别与槽填充两个任务之间的交叉影响,沿着 Liu 等^[28]的思想,Goo 等^[30]提出一种学习意图与语义槽向量之间关系的槽门控机

制(Slot-Gated mechanism),通过问句意图理解将问句转化为单词序列表示并输入到 BiLSTM 网络,门控机制利用上下文向量来建模槽-意图关系,以提高槽填充性能。该结构可以通过全局优化而获得更好的语义框架,越简单的语料语义槽与意图关系更强,更易于建模,因此该方法在较为简单的文本任务中效果优异。在 ATIS 和 Snips 数据集上的实验表明,槽门控机制有效提高了两个任务的交互能力,准确率与使用 Attention 机制相比提高了 4.2%。

Wang 等^[33]提出基于双模型的双向 LSTM 模型的 RNN 语义框架解析网络结构(Bi-model RNN)。其中问句意图理解模块用了两种模型生成词向量:(1)使用前馈模型取一个问句中所有单词向量的平均值作为问句意图;(2)使用递归神经网络将问句中每个词都表示为一个词向量。最后使用消融实验探究解码器对联合模型性能的影响。在涉及多个领域的语料集上的实验表明,带有解码器的联合模型性能更优,在 ATIS 数据集上的意图识别准确率高达 98.99%,比无解码器联合模型的意图识别准确率提升了 0.2%。

在多语言方面,Xu 等^[34]介绍了一种新的多语言自然语言理解语料库 MultiATIS++,它将多语言的 ATIS 语料库扩展到 9 种语言,实验使用一个端到端模型,进行意图识别与槽填充联合任务和跨语言迁移的学习。实验证明该方法在大多数语言上都优于使用快速对齐(fast-align)的简单标签投影方法,仅用一半的训练时间就达到了与其几乎相同的性能,在法语语料库最高取得 97.24% 准确率。

王丽花等^[35]提出了基于门控机制的双向关联模型(BiAss-Gate)来完善现有的联合模型。实验使用 word2vec 模型构建词向量进行问句意图理解,将输入序列的词编码成相对低维稠密的连续向量。基于门控机制的双向关联模型能够利用问句意图的上下文向量和语义槽的上下文向量来建模意图与语义槽之间的关联关系,深入联系了两个任务,建立了双向关联机制,使两个任务的上下文进行交叉融合。实验表明,Bi-Ass-Gate 模型有效弥补了关联模型中浅层或单向关联的不足,在 ATIS 和 Snips 数据集上的意图识别准确率达 98.29%。

华冰涛等^[29]提出一种 BLSTM 结合 CNN 和 CRF 的模型,将问句文本使用 one-hot 编码方法转化成序列编码,再使用双向 LSTM 标注单词标签、CNN 提取句子语义特征、CRF 解码共享表征,最后使用统一损失函数执行意图识别与槽填充任务。实验表明 BLSTM-CNN-CRF 模型在不依赖于人工特征的模型中性能较优,在 ATIS 数据集上的意图识别准确率达到 95.9%。随后,

Chen 等^[31]提出了一个基于 BERT 的意图识别和槽填充联合模型,在大规模的未标记文本数据上,使用 MASK 语言模型和下文预测两种方法对 BERT 模型进行预训练,使用 CRF 提升槽填充预测的准确性。训练后的 BERT 模型提供了依赖于上下文的句子表示,通过微调可用于各种意图分类和槽填充目标任务。实验表明,基于 BERT 的联合模型解决了传统联合模型泛化能力较差的问题。相比于 Slot-Gated 模型,在 ATIS 和 Snips 数据集上的意图识别准确率分别提升 3.8%、1.4%。

总体而言,深度学习模型由于在语义特征提取方面的巨大优势,进一步推动了问句意图理解与识别的发展。模型性能方面,联合模型相比单一模型表现更为优异;数据集上,英文的 ATIS 最受广泛应用,然而,联合模型在中文数据集上的应用尚未发现有较多相关研究,很大原因是标注数据方面的欠缺。此外,现有的深度学习模型都是在规范的问句数据集中进行测试,对于不规范的问句其性能如何仍需后续研究进行探讨。

4 结 语

本文回顾了问句意图理解与识别技术的研究进展,从最初基于规则的方法,发展到基于统计与机器学习方法,截至目前深度学习模型凭借强大的特征提取能力和建模优势,极大推动了问句意图理解与识别的研究,已成为当前该任务的主流方法,并且其中的联合任务模型相对单一模型其性能更高。本节总结一些亟待解决的问题并讨论未来问句意图理解与识别可能的研究方向。

(1) 问句输入不规范。基于意图识别的相关系统与应用针对的目标用户群体有不确定性的特点,这也导致了实际应用中问句意图语义的多样性。在很多情况下,不同用户的问题输入可能含有错别字、口语化、近义词、歧义词、关键词重复、信息缺失和内容宽泛^[3]等非标准语言,使得系统无法精准识别出用户需求。比如用户输入“订票”,这种意图包含了订机票、汽车票、火车票、景点门票和演出门票等;用户输入“大话西游”,可能意图包含电影、电视剧的观看与下载,或者相关音乐、游戏查询等。这些意图不明确的文本要求问句意图理解与识别模型有很高的容错能力与对缺失信息的完善能力。随着问答系统的广泛应用,在线数据集将不断增长,未来可以收集这一类问句构建相应数据集,进而针对性设计例如语义纠错与问句意图理解与识别的联合模型。

(2) 训练样本不均衡。样本不均衡问题对于问句意图理解与识别任务来说是老生常谈的问题。样本(类别)不均衡指的是分类任务中不同类别的训练样例数目差别很大的情况。在样本不均衡的情况下训练模型,会导致样本量少的分类所包含的特征过少^[36],并很难从中提取规律;即使得到分类模型,也容易产生过度依赖于有限的样本数据而导致过拟合问题,当模型应用到新的数据上时,模型的准确性会很差。通常使用的改进方法有两种:一是样本数据上修改,如数据增强;另外可以在训练时修改权重,如 loss 类别加权。

(3) 多意图识别问题。用户输入的问句有时不仅仅包含一个意图,例如“查机票并在机场附近订酒店”,这个查询文本中含有查机票与订酒店两个意图类别;“最近的电影院订票和路线规划”,可能包含最近电影院、订电影票、路线规划这三个意图。此外,很多用户通常是以简短的表达作为问句输入,而短文本通常并不具备足够的信息判断其意图,系统首先要识别出一个问句是否含有多意图,进而再对每个意图进行识别,这种基于短文本的多意图识别问题在目前研究中尚未发现有较好的解决方案。

综上所述,现有研究中问句意图理解与识别系统均以规范长文本、单意图为主^[37],未来的研究将会向多意图、短文本方向发展。随着深度学习的进一步发展,未来期望能够从多样化(不规范、不均衡的样本)的问句中提取有效特征,丰富问句意图特征体系,进一步提升问句意图理解与识别技术的综合性能。

参 考 文 献

- [1] 温丽云. 自然语言理解的联合建模技术研究与应用[D]. 北京:北京邮电大学,2018.
- [2] 张倩. 基于问答系统的问句意图识别方法研究[D]. 重庆:重庆大学,2020.
- [3] 刘娇,李艳玲,林民. 人机对话系统中意图识别方法综述[J]. 计算机工程与应用,2019,55(12):1-7,43.
- [4] Mullick A, Purkayastha S, Goyal P, et al. A framework to generate high-quality datapoints for multiple novel intent detection[EB]. arXiv:2205.02005,2022.
- [5] 郑思露. 基于 ERNIE-BiLSTM-Capsule 的口语意图识别[J]. 数字技术与应用,2022,40(1):128-131.
- [6] Broder A. A taxonomy of web search[J]. ACM SIGIR Forum,2002,36(2):3-10.
- [7] 孙鑫,王厚峰. 问答中的问句意图识别和约束条件分析[J]. 中文信息学报,2017,31(6):132-139.
- [8] 唐素勤,李波,许永敏. 基于句型模板的智能问答系统[J]. 广西师范大学学报(自然科学版),2007,25(2):5-8.

- [9] 曹庆花. 基于领域本体与句型模板的问答系统[D]. 桂林:广西师范大学,2012.
- [10] Kim H, Howland P, Park H, et al. Dimension reduction in text classification with support vector machines[J]. The Journal of Machine Learning Research,2005,6(1):37-53.
- [11] 赵小华. KNN 文本分类中特征词权重算法的研究[D]. 太原:太原理工大学,2010.
- [12] 董云耀,陈小翠,黄炜. 基于 RST 和 SVM 的中文问题分类方法[J]. 杭州电子科技大学学报,2010,30(6):44-47.
- [13] 康海燕,李飞娟,苏文杰. 基于问句表征的 web 智能问答系统[J]. 北京信息科技大学学报(自然科学版),2011,26(1):36-41.
- [14] Li Q, Sun Y C, Xue B. Complex query recognition based on dynamic learning mechanism[J]. Journal of Computational Information Systems,2012,8(20):8333-8340.
- [15] 洪宗祥,李跃新. 基于关系树的知识查询算法研究[J]. 湖北大学学报(自然科学版),2012,34(3):346-349.
- [16] Heck L, Hakkani-Tür D, Tur G. Leveraging knowledge graphs for web-scale unsupervised semantic parsing[C]//Annual Conference of the International Speech Communication Association,2013:1594-1598.
- [17] 林博. 基于用户查询意图识别的个性化搜索方法研究[D]. 湘潭:湘潭大学,2015.
- [18] Chen Y. Convolutional neural network for sentence classification[D]. University of Waterloo,2015.
- [19] Kim Y. Convolutional neural networks for sentence classification[EB]. arXiv:1408.5882,2014.
- [20] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[C]//52nd Annual Meeting of the Association for Computational Linguistics, 2014:655-665.
- [21] Hashemi H B, Asiaee A, Kraft R. Query intent detection using convolutional neural networks[C]//International Conference on Web Search and Data Mining, Workshop on Query Understanding,2016.
- [22] Tang D Y, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification[C]//Conference on Empirical Methods in Natural Language Processing,2015:1422-1432.
- [23] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[EB]. arXiv:1810.04805,2018.
- [24] 迟海洋,严馨,周枫,等. 基于 BERT-BiGRU-Attention 的在线健康社区用户意图识别方法[J]. 河北科技大学学报,2020,41(3):225-232.
- [25] 郑新月,任俊超. 基于 BERT-FNN 的意图识别分类[J]. 计算机与现代化,2021(7):71-76,88.
- [26] Renkens V, Hamme H V. Capsule networks for low resource spoken language understanding[EB]. arXiv:1805.02922, 2018.
- [27] Saha T, Jayashree S R, Saha S, et al. BERT-Caps: A transformer-based capsule network for tweet act classification[J]. IEEE Transactions on Computational Social Systems, 2020,7(5):1168-1179.
- [28] Liu B, Lane I. Attention-based recurrent neural network models for joint intent detection and slot filling[EB]. arXiv:1609.01454,2016.
- [29] 华冰涛,袁志祥,肖维民,等. 基于 BLSTM-CNN-CRF 模型的槽填充与意图识别[J]. 计算机工程与应用,2019,55(9):139-143.
- [30] Goo C W, Gao G, Hsu Y K, et al. Slot-gated modeling for joint slot filling and intent prediction[C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018:753-757.
- [31] Chen Q, Zhuo Z, Wang W. BERT for Joint Intent Classification and Slot Filling[EB]. arXiv:1902.10909,2019.
- [32] Zhang C W, Li Y L, Du N, et al. Joint slot filling and intent detection via capsule neural networks[EB]. arXiv:1812.09471,2018.
- [33] Wang Y, Shen Y L, Jin H X. A bi-model based RNN semantic frame parsing model for intent detection and slot filling[EB]. arXiv:1812.10235,2018.
- [34] Xu W J, Haider B, Mansour S. End-to-end slot alignment and recognition for cross-lingual NLU[EB]. arXiv:2004.14353,2020.
- [35] 王丽花,杨文忠,姚苗,等. 意图识别与语义槽填充的双向关联模型[J]. 计算机工程与应用,2021,57(3):196-202.
- [36] 李晓英,杨名,全睿,等. 基于深度学习的不均衡文本分类方法[J]. 吉林大学学报(工学版),2022,52(8):1889-1896.
- [37] 王浩畅,李斌. 聊天机器人系统研究进展[J]. 计算机应用与软件,2018,35(12):1-6,89.
-
- (上接第8页)
- [46] Rosca S D, Leba M, Sibisanu R C. SSVEP based BCI control of a robot swarm[C]//World Conference on Information Systems and Technologies,2022:296-305.
- [47] Huang D J, Wang X L, Liu J H, et al. Virtual reality safety training using deep EEG-net and physiology data[J]. The Visual Computer,2021,38:1195-1207.
- [48] Zhou Z D, Zhang L W, Wei S, et al. Development and evaluation of BCI for operating VR flight simulator based on desktop VR equipment[J]. Advanced Engineering Informatics,2022,51:101499.
- [49] Tian F, Zhang Y, Li Y J. From 2D to VR film: A research on the load of different cutting rates based on EEG data processing[J]. Information,2021,12(3):130.