



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目：人机对话系统中意图识别方法综述
作者：刘娇，李艳玲，林民
网络首发日期：2019-03-25
引用格式：刘娇，李艳玲，林民. 人机对话系统中意图识别方法综述[J/OL]. 计算机工程与应用. <http://kns.cnki.net/kcms/detail/11.2127.TP.20190322.1815.008.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

人机对话系统中意图识别方法综述

刘 娇, 李艳玲, 林 民

内蒙古师范大学 计算机科学技术学院, 呼和浩特 010022

摘 要: 口语理解是人机对话系统的重要组成部分, 而意图识别是口语理解中的一个子任务, 而且至关重要。意图识别的准确性直接关系到语义槽填充的性能并且有助于后续对话系统的研究。考虑到人机对话系统中意图识别的困难, 传统的机器学习方法无法理解用户话语的深层语义信息, 本文主要对近些年应用在意图识别研究方面的深度学习方法进行分析、比较和总结, 进一步思考如何将深度学习模型应用到多意图识别任务中, 从而推动基于深度神经网络的多意图识别方法的研究。

关键词: 意图识别; 口语理解; 对话系统; 人工智能

文献标志码: A **中图分类号:** TP391.1 **doi:** 10.3778/j.issn.1002-8331.1902-0129

刘娇, 李艳玲, 林民. 人机对话系统中意图识别方法综述. 计算机工程与应用

LIU Jiao, LI Yanling, LIN Min. Review of intent detection methods in human-machine dialogue system. Computer Engineering and Applications

Review of intent detection methods in human-machine dialogue system

LIU Jiao, LI Yanling, LIN Min

College of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010022, China

Abstract: Spoken language understanding(SLU) is a vital part of the human-machine dialogue system, which includes an important sub-task called intent detection. The accuracy of intent detection is directly related to the performance of semantic slot filling, and it is helpful to the following research of the dialogue system. Considering the difficulty of intent detection in human-machine dialogue system, the traditional machine learning methods cannot understand the deep semantic information of user's discourse. This paper mainly analyzes, compares and summarizes the deep learning methods applied in the research of intent detection in recent years, and further considers how to apply deep learning model to multi-intent detection task, so as to promote the research of multi-intent detection methods based on deep neural network.

Key words: intent detection; spoken language understanding; dialogue system; artificial intelligence

基金项目: 国家自然科学基金资助项目(No.61562068, No.11704229, No.61640204, No.61806103); 内蒙古自然科学基金(No.2017MS0607); 内蒙古民委蒙古文信息化专项扶持子项目(No.MW-2014-MGYWXXH-01); 内蒙古自治区“草原英才”工程青年创新创业人才项目; 内蒙古师范大学研究生创新基金(No.CXJJS18112)。

作者简介: 刘娇(1994-), 女, 硕士研究生, 研究领域为自然语言处理, E-mail: 15771337080@163.com; 李艳玲(1978-), 女, 博士, 副教授, 研究领域为自然语言处理, 口语理解, 问答系统; 林民(1969-), 男, 博士, 教授, 研究领域为自然语言处理, 人工智能。

1 引言

随着人工智能时代的到来,越来越多的智能产品已被广泛应用到我们的日常生活中,如情感陪护机器人、私人手机助理 Siri、语音助手 Google Now 和 Cortana、微软亚洲研究院推出的智能聊天机器人小冰以及百度公司推出的小度机器人等。这些智能对话系统不仅可以和用户进行正常的信息交流,而且能为用户的生活带来很多方便。对话系统主要由语音识别 (Automatic Speech Recognition, ASR)、口语理解 (Spoken Language Understanding, SLU)、对话管理 (Dialog Management, DM)、对话生成 (Dialogue Generation, DG) 和语音合成 (Text to Speech, TTS) 这五部分组成^[1],如图 1 所示。为了让机器更好地理解用户的表达,进而为用户反馈正确信息,口语理解起着及其重要的作用。而意图识别 (Intent Detection, ID) 作为口语理解的子模块,也是人机对话系统构成的关键。传统的口语理解主要分为两个子任务——意图识别和语义槽填充。因为早期的研究受应用场景、数据以及计算能力的约束,多数的口语理解限定在某些领域中。但是伴随着技术革新以及多领域对话系统的出现,现今的口语理解时常被分解为三个任务——领域识别,意图识别和语义槽填充^[2]。

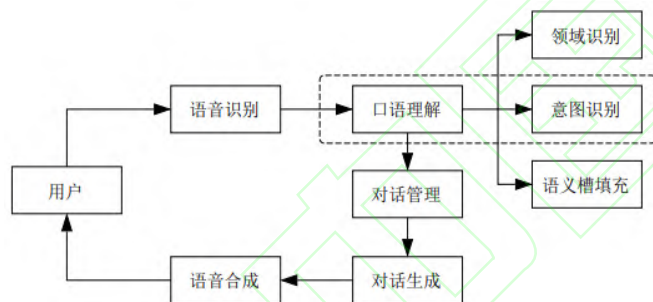


图 1 对话系统示意图

在对话系统中,意图识别至关重要。所谓意图就是用户的意愿,即用户想要做什么。意图有时也被称为“对话行为”(Dialog Act)^[3],即用户在对话中共享的信息状态或上下文变化并不断更新的行为。意图一般以“动词+名词”命名,如查询天气、预订酒店等。而意图识别又称为意图分类,即根据用户话语所涉及到的领域和意图将其分类到先前定义好的意图类别中^[4]。

随着人机对话系统的广泛运用,用户在不同的场合下可能会有不同意图,因而会涉及人机对话系统中的多个领域,其中包括任务型垂直领域和闲聊等。任务型垂直领域的意图文本具有主题鲜明,易于检索的特点,比如,查询机票、天气、酒店等。而聊天类意图文本一般具有主题不明确,语义宽泛、语句简短等

特点,注重在开放域上与人类进行交流。在对话系统中只有明确了用户的话题领域,才能正确分析用户的具体需求,否则会造成后面意图的错误识别。图 2 是口语理解中三个任务应用的实例图,当用户输入一个询问,首先需要明确用户输入的文本所属的话题领域为“火车”还是“航班”,由于意图的类别比话题领域的粒度更细,因此需要根据用户的具体语义信息确定用户的意图是订票、退票还是查询时间,而语义槽的填充也有助于用户意图的判断。所以在人机对话系统的意图识别模块中,首先要对用户话题领域进行识别,接着明确用户的具体意图需求,最终表示成语义框架的形式。

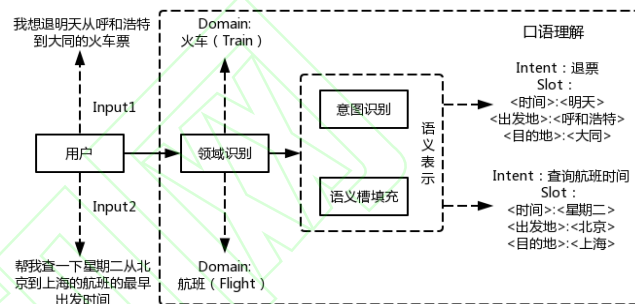


图 2 意图识别实例图

2 意图识别的难点

2.1 数据来源匮乏

随着人工智能技术的不断发展,大型互联网公司推出聊天机器人,由于用户体验性较低,大多数研究者难以获取到用户与机器人之间的聊天文本,因而导致需要研究的对话文本数量有限,这成为意图识别任务面临的重大问题^[5]。在实际的意图识别过程中,带标注的意图文本特别少,获取也十分困难,这也给意图识别的研究和发展带来了挑战^[6]。

2.2 用户表达的不规范性

在聊天系统中,用户的意图表达文本一般具有表达口语化、语句简短、内容宽泛等特点,这就使得意图识别较为困难。例如,“我想找个吃饭的地方”,这种口语化的日常用语表达对应的意图则是“找餐馆”,所以意图文本的口语化使得领域主题不明确,不利于用户意图的识别。对于“汉庭”这种语义不丰富的意图表达方式,虽然“汉庭”常常与“酒店”搭配出现,但让机器识别出用户的话题领域为“酒店”,则是一项非常困难的任务。对于“我想订票”这种意图表达方式而言,订票有可能是订机票、订火车票、订汽车票等。由于用户的意图表达内容太笼统,导致机器不能及时为用户反馈结果。

2.3 意图的隐含性

随着人机对话系统应用范围的不断扩大,意图的表达方式也越来越多。有些意图表达很明确,而有些

意图表达却蕴含有更深层的意思。因此意图按照表达种类可以分为显式意图和隐式意图^[7]，显式意图指用户通过文本形式，明确指出自己的意图需求，包含话题领域、意图类别等内容。隐式意图指用户没有明确自己的意图需求，需要通过分析用户的潜在意图，来推理用户的真实意图^[8]，用户常用的意图表达文本如表 1 所示。

表 1 用户常用的意图表达

序号	显式意图文本	隐式意图文本
1	我想订一张去北京的机票 (Flight)	明天我要去北京参加会议 (Flight)
2	预订人民公园附近的酒店一晚 (Hotel)	下周我要去深圳玩两天 (Hotel)
3	我想买一部手机 (Shopping)	iPhone 8 好用吗 (Shopping)

从表 1 的三个例子可以看出，每个例子的两种文本表达方式截然不同，但具有相同的意图。例 1 的显式意图文本明确用户需要订机票，而隐式意图文本虽然没有相关的领域信息，通过推测，用户可能需要订去北京的机票。例 2 的显式意图文本表明用户想要预定酒店，隐式意图文本没有直接明确意图需求，通过推测表明用户需要预订深圳的酒店。例 3 的显式意图文本明确用户具有消费意图，而隐式意图文本没有直接说明购买意图，但是通过分析，表明用户具有潜在的消费意图。所以没有明确话题领域和类别信息的隐式意图识别也成为意图识别中的一大难点。

2.4 意图的多样性

用户表达的一句话中有时不是仅含有一种意图，而是含有多种意图。如：“我想预订一张从呼和浩特到上海的机票，并预订当天上海外滩周围的酒店一晚”，这句话同时包含订机票和预定酒店两种意图。在用户表达的一句话中同时识别出两种甚至多种意图的问题成为多意图识别问题。多意图识别类似于多标签 (Multi-label, ML) 分类，但又不同于多标签分类，多标签分类通常处理的是长文本，而多意图识别主要针对短文本进行处理。如何在较短的文本内识别出用户的多种意图则是意图识别的又一个难点。在多意图识别过程中，首先需要分析用户意图文本是否包含多种意图，如果用户意图文本包含多种意图，如何准确识别用户的多种意图是值得我们思考的问题。

3 传统的意图识别方法

近年来，意图识别成为学术界和工业界新的研究热点，为了正确理解人机对话系统中的用户意图，

大部分学者将意图识别看作是一种语义话语分类 (Semantic Utterance Classification, SUC) 问题^[9]。传统的意图识别方法主要有基于规则 (Rule-Based) 模板的语义识别方法^[10] (1993) 和使用统计特征的分类算法^[11,12] (2002-2014)。

3.1 单意图识别方法

3.1.1 基于规则模板的意图识别方法

基于规则模板的意图识别方法一般需要人为构建规则模板以及类别信息对用户意图文本进行分类^[13]。Ramanand 等人^[14] (2010) 针对消费意图识别，提出基于规则和图的方法来获取意图模板，在单一领域取得了较好的分类效果。Li 等人^[15] (2015) 研究发现在同一领域下，不同的表达方式会导致规则模板数量的增加，需要耗费大量的人力物力。所以，基于规则模板匹配的方法虽然不需要大量的训练数据，就可以保证识别的准确性，但是却无法解决意图文本更换类别时带来重新构造模板的高成本问题。

3.1.2 基于统计特征分类的意图识别方法

基于统计特征分类的方法，则需要对语料文本进行关键特征的提取，如字、词特征、N-Gram 等，然后通过训练分类器实现意图分类。常用的方法有朴素贝叶斯 (Naive Bayes, NB)^[16] (1998)、AdaBoost^[17] (2000)、支持向量机 (Support vector machine, SVM)^[18] (2003) 和逻辑回归^[19] (2007) 等。

陈^[20] (2014) 分别使用 SVM 和 Naive Bayes 分类器对微博语料进行消费意图分类，F1 值都达到 70% 以上，但这两种分类器都需要人工提取特征，不仅成本高，而且特征的准确性无法得到保障，同时还会导致数据稀疏问题。由于 SVM 对多类别数据信息的分类效果不好而且泛化性能较差，贾^[21] (2016) 通过引入 AdaBoost 算法和 PSO 算法，利用 POS 优化 SVM 参数，并且用 AdaBoost 算法集成 PSOSVM 分类器，得到一种 AdaBoost-PSOSVM 强分类器，在相同数据集上分类性能明显高于 SVM 分类器。但这些方法都不能准确理解用户文本的深层次语义信息。

3.2 多意图识别方法

对于多意图识别任务，Kim 等人^[22] (2017) 提出一种基于单意图标记训练数据的多意图识别系统。他将句子看作三种类型，单意图语句、带连词的多意图语句和不带连词的多意图语句，然后采用两阶段法实现多意图识别，如图 3 所示。该研究将用户意图文本含有的意图数量最多限制为两种，如果第一阶段识别出的意图少于两种，则执行第二阶段，如果第二阶段识别出的意图同样少于两种，则进行单意图识别。第一阶段，该系统根据输入句子中的连词生成多意图假

设集合 H ，其中一个多意图假设 $h \in H$ ，可以表示为 $\langle h_{left}, h_{conj}, h_{right} \rangle$ ， h_{left} 表示连词左侧的子句， h_{conj} 表示连接词， h_{right} 表示连词右侧的子句。然后采用最大熵模型^[23]对假设进行评估，选取满足特定条件的最优假设。第二阶段，系统对输入的句子进行顺序标注和意图标记，采用线性链条件随机场（Conditional Random Fields, CRF）分类器^[24]进行意图识别。实验结果表明采用两阶段法实现多意图识别性能优于单阶段方法。该方法在多意图识别任务上取得了不错的进展。

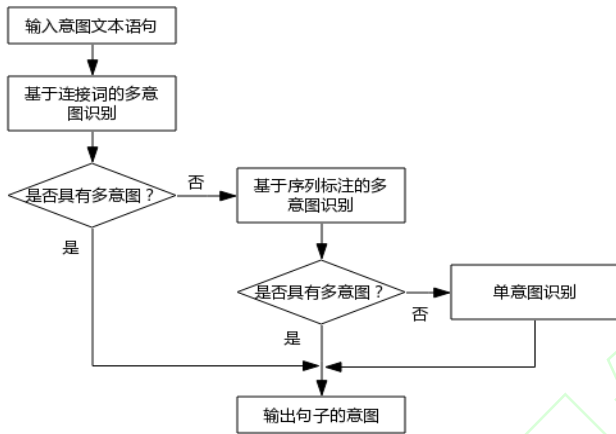


图 3 多意图识别流程图

4 基于深度学习的意图识别方法

随着深度学习的不断发展，越来越多的学者们将词向量、卷积神经网络（Convolutional neural networks, CNN）、循环神经网络（Recurrent Neural Network, RNN）和其变体长短时记忆（Long Short-Term Memory, LSTM）网络、门控循环单元（Gated Recurrent Unit, GRU）、注意力机制和胶囊网络应用于意图识别任务中，相比于传统的机器学习方法，深度学习模型在识别性能上有了很大的提升。

4.1 单意图识别方法

4.1.1 基于词向量的意图识别

近年来，在自然语言处理过程中，由于使用原始词法特征会导致数据稀疏问题，词向量逐渐被用于语义分析任务中，而且连续表示学习可以解决数据稀疏问题^[25]（2003）。Kim 等人^[26]（2015）将词向量作为词法特征进行意图分类，与传统的词袋模型相比，基于词向量的意图分类方法对不同分类内容的表征能力和领域可扩展性更好。

考虑到词向量的语义信息不全等问题，Kim 等人^[27]（2016）利用语义词汇字典（如 WordNet^[28]和 Paraphrase Database（PPDB）^[29]）的信息来丰富词向量，从而提高意图文本的语义表示，通过构建 BLSTM

（Bidirectional Long Short-Term Memory）模型进行意图识别。在航空旅行信息系统（Air Travel Information System, ATIS）数据集和来自 Microsoft Cortana 的关于地点的真实日志数据集上验证，表明丰富的语义词汇向量可以提高意图的识别性能。而且对于规模较小的训练集采用复杂的深度学习模型，提供丰富的词向量会对模型性能有一定的帮助。所以词向量的研究会对深度学习模型的运用起到至关重要的作用。

4.1.2 基于卷积神经网络的意图识别

CNN 最初被用于图像处理^[30]，随着词向量技术的出现，CNN 被广泛应用于自然语言处理领域^[31]，并且取得了很好的研究成果。Kim^[32]（2014）等人尝试将 CNN 用于文本分类任务中，并取得了十分理想的效果。基于此，Hashemi 等人^[33]（2016）采用 CNN 提取文本向量表示作为查询分类特征来识别用户搜索查询的意图，与传统的人工特征提取方法相比，不仅减少了大量的特征工程，而且可以得到更深层次的特征表示。但是 CNN 只能提取到意图文本的局部语义特征，不能保持语义的连贯性。

4.1.3 基于循环神经网络及其变体的意图识别

RNN 不同于 CNN，它表示的是一个词序列，而且可以根据上下文学习词序语义信息。Bhargava A^[34]（2013）将上下文信息纳入意图识别任务中降低了意图识别的错误率，说明上下文信息有助于意图的识别。一个简单的 RNN 存在梯度爆炸或梯度消失等问题，不能很好地模拟长期依赖关系。

而序列建模最流行的网络 LSTM^[35]（1997）通过在 RNN 结构中引入一个内存单元来解决这个问题，同时它可以控制要保留和遗忘的信息。该模型也常被用于解决意图识别问题，Ravuri 等人^[36]（2015）提出用 RNN 和 LSTM 两种模型来解决意图分类问题，将两种模型分别在 ATIS 数据集上进行实验，结果表明 LSTM 模型的意图识别错误率比 RNN 低 1.48%。主要是因为 LSTM 对文本的时序关系具有良好的建模能力，而且对输入较长的文本具有很好的记忆功能。

GRU 是 LSTM 模型的一种改进^[37,38]，具有在长序列上保留信息的能力，而且可以学习上下文语义信息。它和 LSTM 在大部分实验中都优于 RNN^[36,39]，相比于 LSTM，GRU 只使用两个门即重置门（reset gate）和更新门（update gate），模型结构更简单，含有的参数更少，需要的文本语料更少^[38]。而双向门控循环单元（Bidirectional Gated Recurrent Unit, BGRU）可以充分考虑上下文语义信息从而对意图文本进行更好的特征表示，通常将隐藏状态的最终输出作为意图文本表示，从而得到意图类别结果。

针对意图识别任务, Ravuri 等人^[39] (2016) 又采用图 4 所示的 GRU 和 LSTM 在 ATIS 和 Cortana 数据集上进行全面比较。实验表明 GRU 和 LSTM 模型在意图分类任务上的性能几乎一样, 但是 GRU 的参数更少, 模型更简单。

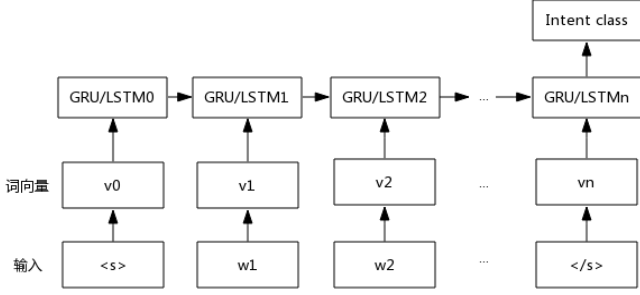


图 4 LSTM 和 GRU 模型用于意图分类

4.1.4 深度学习模型的组合运用于意图识别

考虑到各种深度学习模型的优缺点, 大部分研究者将具有不同优势的深度学习模型进行组合对用户意图进行分类。钱^[8] (2017) 利用 CNN 可以更深层的提取意图文本特征以及 LSTM 可以对文本的时序关系建模的优点, 提出基于 Convolutional-LSTM 的出行消费意图识别模型, 并且取得了很好的性能。余等人^[40] (2018) 针对短文本会导致数据稀疏的问题, 提出了基于短文本主题模型 (Biterm Topic Model, BTM) 和 BGRU 的多轮对话意图识别模型, 该混合模型在用户就医意图识别上取得了很好的效果, 而且优于文献 [8] 的性能。黄^[41] (2018) 提出了 Character-CNN-BGRU 深度学习组合模型, 该组合模型利用基于字符的方法不仅使得所用词表范围更小而且可以解决未登录词问题的优势, 再结合 CNN 可以提取到意图文本的深层局部特征以及 BGRU 可以保证文本的时序关系对意图识别任务进行建模, 突出组合模型在意图识别任务上的优势。但是组合模型结构复杂, 训练时间较长, 如何简化组合模型是值得思考的问题。

4.1.5 基于 BLSTM 的自注意力 (self-attention) 模型的意图识别

词向量已经被广泛应用于自然语言处理领域中, 随着深度学习模型的发展, 出现了各种句子级别向量的表示, 如: 使用 CNN 中的最大池化 (max-pooling) 或平均池化 (mean-pooling) 得到句子向量, 使用 RNN 的隐藏状态或最终隐藏状态创建句子表示等。Lin 等人^[42] (2017) 提出了一种通过引入自注意力机制提取句子表示的模型改进原有方法, 通过二维矩阵来表示句子向量, 将句子的不同语义信息用多个向量进行表示。该模型在 BLSTM 上执行, 通过对 BLSTM 的隐层状态加权求和得到句子向量表示, 从而实现意图分类。该模型通过自注意力机制可以获取到句子的多种

语义信息, 有助于多意图识别的研究。

图 5 为自注意力模型图, 如图 5(a) 所示, 假设输入序列为 $S = (w_1, w_2, \dots, w_n) \in R^{n \times d}$, 将 S 输入到 BLSTM 中, 得到第 t 个单词的前后项隐藏状态计算如下:

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

$$\overleftarrow{h}_t = \overleftarrow{LSTM}(w_t, \overleftarrow{h}_{t+1}) \quad (2)$$

如果 LSTM 的隐藏单元数为 k , 将前后项隐藏状态连接得到 h_t , 且 $h_t \in R^{2k}$, 则输入的意图文本可以用 $H \in R^{n \times 2k}$ 表示所有隐藏状态 h_t 的集合, 即 $H = (h_1, h_2, \dots, h_n)$ 。

将自注意力权重矩阵表示为:

$$a = \text{soft max}(W_{s2} \tanh(W_{s1} H^T)) \quad (3)$$

如图 5(b) 所示, $W_{s1} \in R^{m \times 2k}$ 和 $W_{s2} \in R^{1 \times m}$ 为自注意力权重矩阵, m 为自注意力的隐藏单元数, 是一个超参数, 可以任意设置, $H^T \in R^{2k \times n}$ 为 H 的转置矩阵。通过计算可以得到 $a \in R^{1 \times n}$ 。所以意图文本向量可以表示为: $d = a \cdot H$, 则 $d \in R^{1 \times 2k}$ 。

如果从意图文本中提取 r 个语义特征, 则需要 r 个自注意力头部进行语义特征提取, 则 $W_{s2} \in R^{r \times m}$, 得到 $A \in R^{r \times n}$, 最终的意图文本向量可以表示为: $D = A \cdot H$, 则 $D \in R^{r \times 2k}$ 。

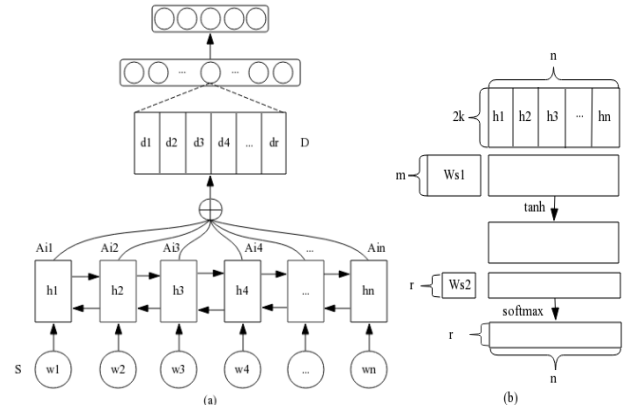


图 5 自注意力模型图

4.1.6 胶囊网络模型用于意图识别

“胶囊”的概念最先由 Hinton 等人^[43] (2011) 提出, 用来解决 CNN 的表征局限性, 一个胶囊包含一组神经元的向量表示, 向量的方向表示实体属性, 向

量的长度表示实体存在的概率。Sabour 等人^[44](2017)提出了胶囊网络,将 CNN 的标量输出特征检测器用矢量输出胶囊代替,并且通过协议路由代替最大池化。相比于原来的 CNN,胶囊网络会保持实体在区域内的准确位置信息。因此,Zhao 等人^[45](2018)首次将胶囊网络用于文本分类任务,提出三种动态路由策略来提高动态路由过程的性能,以减轻噪声(停用词和与类别无关的词)胶囊的干扰。在标准数据集上的实验,表明胶囊网络在文本分类任务上具有很好的性能,同时在多标签文本分类任务上也表现出很好的效果。

在意图识别任务上,Xia 等人^[46](2018)提出了一个基于胶囊网络的意图胶囊模型,该模型利用胶囊模型在文本建模中的优势对文本进行分层建模,如图 6 所示,首先采用文献[42]的方法对意图文本提取具有自注意力的语义特征,即语义向量。由于不同用户对同一种意图的表达式不同,但对一种意图的贡献大于其他意图,通过采用动态路由机制动态分配每种语义的适当贡献度将其聚合形成更高级别的预测向量,即意图语义表示,从而进行意图分类。该模型在意图识别任务上取得了不错的效果。

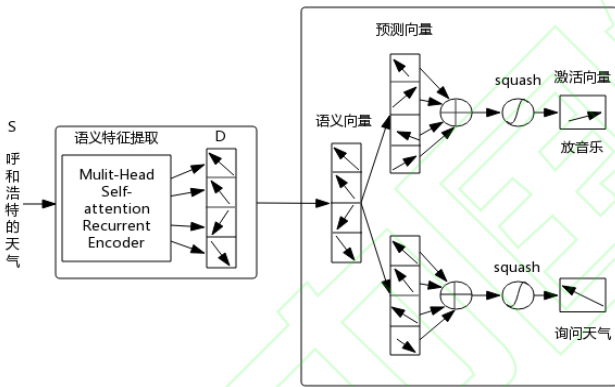


图 6 基于意图胶囊的意图识别过程图

4.1.7 联合识别的方法用于意图识别

随着意图识别方法的不断研究与改进,考虑单一任务研究因其独立的模型而容易出现错误传播,有些学者提出了语义槽填充和意图识别的联合模型。李等人^[47](2017)通过三角链条件随机场对意图识别和语义槽填充进行联合建模,与将语义槽填充结果作为意图识别特征的级联模型相比,联合模型在意图识别任务上表现出很好的性能,突出了两者之间的关联性。Liu 等人^[48](2016)通过在双向循环神经网络的隐层上增加注意力机制来捕捉句子的重要语义成分,从而提高意图识别的准确率,在 ATIS 数据集上验证基于注意力机制的双向循环神经网络模型的意图识别错误率为 2.35%。研究者通过将语义槽填充与意图识别联合实验,在相同数据集上验证,意图识别的错误率降到 1.79%。可见,语义槽填充有助于意图识别。

4.2 多意图识别方法

针对多意图识别,杨等人^[49](2018)将用户意图文本进行依存句法分析确定是否包含多意图,利用词频-逆文档频率(TF-IDF)和训练好的词向量计算矩阵距离来确定句子的意图数量,通过将句法特征和 CNN 结合进行意图分类,最终判别用户的多种意图。该方法在 10 种类别的多意图识别任务中取得了不错的效果。到目前为止,基于深度神经网络的多意图识别研究较少,所以这是值得我们研究的一个方向。

5 意图识别的评价方法

5.1 单意图识别的评价方法

目前,意图识别普遍被看作是语义话语分类问题,因此采用文本分类器中常用的评价指标对意图识别方法的性能进行评价^[40]。即准确率、召回率、F1 值和错误率、分类速度等。

(1) 准确率 (Accuracy)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

(2) 精确率 (Precision)

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

(3) 召回率 (Recall)

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

(4) F1 值 (F1-score)

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

其中, TP (True Positive) 表示真阳性,即类 A 的样本被正确分类到类 A 中, TN (True Negative) 表示真阴性,即不属于类 A 的样本被正确分类到类 A 以外的其他类, FP (False Positive) 表示假阳性,即不属于类 A 的样本被错误分到类 A 中, FN (False Negative) 表示假阴性,即属于类 A 的样本被错误分到类 A 以外的其他类中,如表 2 所示。

表 2 TP、FN、FP、TN 的具体说明

	A 类样本	其他类样本
在 A 类中检索到	TP (True Positive)	FP (False Positive)
在其它类中检索到	FN (False Negative)	TN (True Negative)

5.2 多意图识别的评价方法

多意图识别方法比单意图识别方法复杂,因此多意图识别的评价方式也不同于单意图识别评价方式,

在评价时需要对句子中意图预测的类别都进行判断。采用文献[49]的评价指标。假设有 N 个多意图样本记为 (x_i, A_i) , $0 \leq i \leq N$ 。 A_i 表示样本集正确意图结果的集合。 $B_i = S(x_i)$ 表示待测样本集的预测意图结果的集合。具体计算公式如下:

(1) 准确率 (Accuracy)

$$Accuracy = \frac{1}{N} \sum_{i=1}^N \frac{|A_i \cap B_i|}{|A_i \cup B_i|} \quad (8)$$

(2) 精确率 (Precision)

$$Precision = \frac{1}{N} \sum_{i=1}^N \frac{|A_i \cap B_i|}{|B_i|} \quad (9)$$

(3) 召回率 (Recall)

$$Recall = \frac{1}{N} \sum_{i=1}^N \frac{|A_i \cap B_i|}{|A_i|} \quad (10)$$

6 总结与展望

本文主要对人机对话系统中的意图识别的难点和方法进行介绍。主要针对深度学习模型的意图识别方法进行总结对比,传统的意图识别方法不能深层次地理解用户的意图,而深度学习模型则显示出了它的优势。传统的对话系统主要面向特定领域的单意图识别,目前单意图识别已被大部分学者研究。随着人类与机器的交互越来越频繁,用户的话语表达不仅仅只限定于一种意图,如何准确识别出用户的多种意图目前的研究比较少,所以我们下一步主要围绕以下两点对基于深度学习的多意图识别进行研究:

(1) 在基于多头自注意力模型的意图识别过程中,自注意力模型可以利用多个自注意力头部对句子的多种语义特征进行提取,从而有助于多意图识别的研究。如何运用自注意力模型识别用户文本中包含的多种意图值得我们思考。

(2) 胶囊网络模型在意图识别任务上获得了很好的性能,而且对多标签分类也有很好的效果,而多意图识别又类似于多标签分类,所以胶囊网络模型的研究为基于深度学习的多意图识别研究奠定了一定的基础,如何运用胶囊网络模型识别用户文本的多种意图是我们下一步的主要工作。

参考文献:

- [1] Chen Hongshen, Liu Xiaorui, Yin Dawei, et al. A survey on dialogue systems: recent advances and new frontiers [J]. SIGKDD Explorations, 2017, 19(2):25-35.
- [2] Tur G. Spoken Language Understanding: Systems for extracting semantic information from speech [D]. New York, NY: John Wiley and Sons, 2011.
- [3] Austin J A. How to do things with words [M]. Harvard University Press. Cambridge: 1962.
- [4] Celikyilmaz A, Hakkani-Tur D, Tur G, et al. Exploiting distance based similarity in topic models for user intent detection [C]// Automatic Speech Recognition & Understanding. IEEE, 2011: 425-430.
- [5] Luo Bingfeng, Feng Yansong, Wang Zheng, et al. Marrying up Regular Expressions with Neural Networks: A case study for spoken language understanding [C]// Proc of the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 2083-2093.
- [6] 李艳玲, 颜永红. 中文口语理解弱监督训练方法[J]. 计算机应用, 2015, 35(7):1965-1968.
- [7] Chen Zhiyuan, Liu Bing, Hsu Meichun, et al. Identifying intention posts in discussion forums [C]// Proc of Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies. Atlanta, 2013:1041-1050.
- [8] 钱岳. 聊天机器人中用户出行消费意图识别方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2017.
- [9] Dauphin Y N, Tur G, Hakkani-Tur D, et al. Zero-Shot learning for semantic utterance classification [EB/OL]. (2013)[2018-12-25]. <https://arxiv.org/abs/1401.0509v3>.
- [10] Appelt D, Bear J, Cherny L, et al. GEMINI: A natural language system for spoken-language understanding [C]// Proc Meeting of the Association for Computational Linguistics. 1993:54-61.
- [11] 燕鹏举.对话系统中的自然语言理解研究 [D], 北京: 清华大学,2002.
- [12] Ahmad A S, Hassan M Y, Abdullah M P, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting [J]. Renewable & Sustainable Energy Reviews, 2014, 33(2):102-109.
- [13] Ramanand J, Bhavsar R K, Pedaneka R N. Wishful thinking: finding suggestions and 'buy'wishes from product reviews[C]// CAAGET '10: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Stroudsburg, PA: Association for Computational Linguistics, 2010: 54-61.
- [14] Prager J, Radev D, Brown E, et al. The Use of Predictive Annotation for Question Answering in TREC8[C]// Conference of the Eighth Text Retrieval, 1999:399-411.
- [15] Li X, Dan R. Learning question classifiers: the role of semantic information[J]. Natural Language Engineering, 2015, 12(3):229-249.
- [16] Andrew McCallum, Kamal Nigam, et al. A comparison of event models for naive bayes text classification [C]// In AAAI-98 workshop on learning for text categorization. 1998:41-48.
- [17] Schapire R E, Singer Y. BoosTexter: a boosting-based system for text categorization [J]. Machine Learning, 2000, 39(2-3):135-168.
- [18] Haffner P, Tur G, Wright J H. Optimizing SVMs for complex call classification [C]// IEEE International Conference on Acoustics. IEEE, 2003:632-635.
- [19] Genkin A, Lewis D D, Madigan D. Large-Scale Bayesian Logistic Regression for Text Categorization [J]. Technometrics, 2007, 49(3):291-304.
- [20] 陈浩辰.基于微博的消费意图挖掘[D].哈尔滨: 哈尔滨工业大学, 2014.
- [21] 贾俊华.一种基于 AdaBoost 和 SVM 的短文本分类模型 [D].河北: 河北工业大学, 2016.
- [22] Kim B, Ryu S, Gary G L. Two-stage multi-intent detection for spoken language understanding [J]. Multimedia Tools and Applications, 2017, 76(9):1-14.
- [23] Ratnaparkhi A. Maximum Entropy Models For Natural

- Language Ambiguity Resolution [M]. University of Pennsylvania, 1998.
- [24] Lafferty J, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [J]. *Proceedings of Icml*, 2001, 3(2):282-289.
- [25] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model [J]. *Journal of Machine Learning Research*, 2003, 3 (2): 1137-1155.
- [26] Kim D, Lee Y, Zhang J, et al. Lexical feature embedding for classifying dialogue acts on Korean conversations [C]// *Proc of 42th Winter Conference on Korean Institute of Information Scientists and Engineers*, 2015: 575-577.
- [27] Kim J K, Tur G, Celikyilmaz A, et al. Intent detection using semantically enriched word embeddings [C]// *Spoken Language Technology Workshop*. IEEE, 2016:414-419.
- [28] Fellbaum C, Miller G. WordNet: An electronic lexical database [J]. *Library Quarterly Information Community Policy*, 1998, 25(2):292-296.
- [29] Pavlick E, Rastogi P, Ganitkevitch J, et al. PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification [C]// *Meeting of the Association for Computational Linguistics & the International Joint Conference on Natural Language Processing*. 2015:425-430.
- [30] Lecun Y L, Bottou L, Bengio Y, et al. Gradient-Based Learning Applied to Document Recognition [J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.
- [31] Peng Wang, Xu Jiaming, Bo Xu, et al. Semantic Clustering and Convolutional Neural Network for Short Text Categorization [C]// *Proceedings ACL*, 2015:352-357.
- [32] Kim Y. Convolutional Neural Networks for Sentence Classification [C]// *Proc of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014:1746-1751.
- [33] 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.
- [34] Bhargava A, Celikyilmaz A, Hakkanitur D, et al. Easy Contextual Intent Prediction and Slot Detection [C]// *IEEE International Conference on Acoustics*. IEEE, 2013: 8337-8341.
- [35] Hochreiter S, Schmidhuber J. Long Short-Term Memory [J]. *Neural Computation*, 1997, 9(8):1735-1780.
- [36] Ravuri S V, Stolcke A. Recurrent neural network and LSTM models for lexical utterance classification [C]// *16th Annual Conference of the International Speech Communication Association*. 2015:135-139.
- [37] Dey R, Salemt F M. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks [C]// *IEEE 60th International Midwest Symposium on Circuits and Systems*. IEEE, 2017:1597-1600.
- [38] Chung J, Gulcehre C, Cho K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks On Sequence Modeling[EB/OL].(2014)[2018-12-25]. <https://arxiv.org/abs/1412.3555>.
- [39] Ravuri S, Stolcke A. A comparative study of recurrent neural network models for lexical domain classification [C]// *Proc of the 41th IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2016: 6075-6079.
- [40] 余慧, 冯旭鹏, 刘利军, 等. 聊天机器人中用户就医意图识别方法 [J]. *计算机应用*, 2018, 38(8): 2170-2174.
- [41] 黄佳伟. 人机对话系统中用户意图分类方法研究 [D]. 武汉: 华中师范大学, 2018.
- [42] Lin Zhouhan, Feng Minwei, Santos C N D, et al. A structured self-attentive sentence embedding [C/OL]. (2017)[2018-12-25]. <https://arxiv.org/pdf/1703.03130.pdf>.
- [43] Hinton G E, Krizhevsky A, Wang S D. Transforming Auto-Encoders [C]// *In International Conference on Artificial Neural Networks*, 2011:44-51.
- [44] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules [C]// *In Advances in Neural Information Processing Systems*. 2017:3859-3869.
- [45] Zhao Wei, Ye Jianbo, Yang Min, et al. Investigating Capsule Networks with Dynamic Routing for Text Classification [C]// *Proc of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018:3110-3119.
- [46] Xia Congying, Zhang Chenwei, Yan Chenwei, et al. Zero-shot User Intent Detection via Capsule Neural Networks [C]// *Proc of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018:3090-3099.
- [47] 李艳玲, 颜永红. 统计中文口语理解执行策略的研究 [J]. *计算机科学与探索*, 2017,11(6): 980-987.
- [48] Liu B, Lane I. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling [C]// *17th Annual Conference of the International Speech Communication Association*. 2016:685-689.
- [49] 杨春妮, 冯朝胜. 结合句法特征和卷积神经网络的多意图识别模型[J]. *计算机应用*, 2018, 38(7): 1839-1845.