

任务型与问答型对话系统中的语言理解技术

关键词：对话系统 语言理解

车万翔 张 宇
哈尔滨工业大学

语言理解

在各种类型的对话系统中，私人助理是一种非常重要的类型。它是通过自然语言对话的方式，完成用户发出的指令类任务，或者回答一些常识知识，例如，天气查询、机票预定、知识问答等。有很多著名的科技公司涉猎这一类型的对话系统，如苹果Siri，微软小娜，谷歌助理(Google Assistant)，百度度秘，讯飞AIUI等。

一个对话系统通常有三个关键模块，即语言理解(Language Understanding)、对话管理(Dialogue Management)和语言生成(Language Generation)。其中语言理解技术指的是将用户面向机器说出的自然语言转换为语义表示^[1]，随后对话管理模块根据语义表示、语义上下文、用户元信息等，找到合适的执行动作，再根据具体的动作生成一句自然语言来回复用户^[2]。

任务型对话系统的语言理解模块通常可以拆分为领域识别、意图识别、语义槽¹(slot)填充等三个子任务。例如，对于用户输入的语句：

查一下明天北京到上海的机票。

表1列出了这句话的领域、意图、语义槽，其中，“明

表1 领域、意图、语义槽示例

句子	查一下	明天	北京	到	上海	的	机票
领域	航旅						
意图	查询机票						
语义槽		flight_date	from_city		to_city		

天”的语义槽类型是出发时间，“北京”和“上海”的语义槽类型分别是出发地和到达地。

知识问答型对话系统的语言理解模块通常指的是对问题进行分类，即为问题返回一个类别标签，该类别标签是问题期望答案的类别，如时间、地点等。

领域/意图与问题分类

任务型对话中的领域识别和意图识别，以及知识问答型对话系统中的问题分类通常都可以视为文本分类任务来解决。

其中，领域和意图的定义与对话系统自身的定位和所具有的知识库有很大关系，即具有非常强的领域相关性。而问题分类的类别通常相对通用，具有一定的领域无关性，如哈尔滨工业大学信息检索和社会计算研究中心提出的中文问题的分类体系^[3]，主要包括

¹ 语义槽相当于意图中所涉及的关键信息，包含语义槽名称和语义槽值。语义槽名称如“出发地”，语义槽值为对应的具体实体内容，如“北京”。语义解析后，返回语义槽名称和语义槽值。

表2 中文问题分类体系

大类	小类
描述(DES)	简写、表达、意思、方式、原因、定义、判断、其他
人物(HUM)	特定人物、机构团体、人物描述、其他
地点(LOC)	宇宙、城市、大陆、国家、省、河流、湖泊、山脉、大洋、岛屿、建筑、地址、其他
数字(NUM)	温度、面积、体积、重量、速度、频率、距离、钱数、数量、顺序、倍数、百分比、号码、时间长度、范围、其他
时间(TIME)	年、月、日、季节、时代、星期、节气、节假日、时间、时间范围、其他
实体(OBJ)	物质、动物、植物、微生物、身体、材料、机具、衣物、食物药品、货币、票据、语言、事件、疾病、艺术品、服务、文字作品、学术学科、计划规划、法律法规、职位头衔、职业行业、符号、奖励、刑法、类别、权利义务、颜色、宗教、运动娱乐、术语、其他

描述类 (DES)、人物类 (HUM)、地点类 (LOC)、数字类 (NUM)、时间类 (TIME)、实体类 (OBJ) 6 个大类, 每个大类又划分为若干小类, 共计 84 个小类(见表 2)。

国内外学者已经对文本分类技术进行了大量的研究, 基于深度学习的方法依然是目前的主流。Kalchbrenner 等人^[4]构建了一种动态卷积神经网络(DCNN), 对句子语义信息进行建模从而对文本进行分类, 在 DCNN 中使用了一种全局 k-max 池化操作方法。Mou 等人^[5]利用依存句法树和成分树构建了一种基于树的卷积神经网络(TBCNN)。TBCNN 利用最大池化方法融合了句子的结构特征及其他多种特征。Komninos 等人^[6]在神经网络模型中使用基于上下文的词嵌入特征。基于深度学习的方法可以自动抽取特征, 在减少了工作量的同时, 还可以抓取到直觉上无法获取的特征, 从而提高分类的精度。

语义槽填充

任务介绍

语义槽填充任务的主要目标是在已知特定领域或意图的语义框架 (semantic frame) 的前提下, 从输入语句中抽取该语义框架中预先定义好的语义槽的值。语义槽填充任务可以转化为序列标注任务, 即使用经典的 IOB² 标记法。用此标注法, “查一下明天北京到上海的机票”, 标注结果为:

句子	查一下	明天	北京	到	上海	的	机票
语义槽标签	O	B-from_date	B-from_city	O	B-to_city	O	O

研究现状

基于特征抽取的传统机器学习方法

基于特征抽取的传统机器学习模型已经在语义槽填充任务上大量应用。例如, 基于隐马尔科夫模型 (HMM) 或上下文无关文法 (CFG)^[7-9], 隐向量状态 (HVS)^[10] 等的生成模型, 以及基于条件随机场 (CRF)^[11-17], 支持向量机 (SVM)^[18] 等的判别模型。

基于深度学习的方法

随着深度学习技术在自然语言领域的发展, 基于深度学习的方法也广泛应用于语义槽填充任务。相比于传统的机器学习方法, 深度学习模型不仅能够自动学习输入数据的隐含特征, 同时也能够融入人工抽取的特征。

Yao 等人^[19]和 Mensil 等人^[20]最早将循环神经网络 (RNN) 应用于语义槽填充的任务上。之后双方总结了一篇使用 RNN 进行语言理解任务中语义槽填充的文章^[2]。文章将语义槽填充视为序列标注任务, 对不同种类 RNN、双向 RNN、RNN-CRF 等网络结构的表现做了详细实验, 结果表明, 相对于 CRF 的基线方法, 这些模型结构能够获得更好的结果。Yao 等人使用了 LSTM 网络及相关变种^[21], 还有学者使用了 CNN 网络^[22-24], Chen 等人^[25, 26]使用了 Encoder-Decoder 的框架进行语义槽标签标注, 并引入了注意力机制。

意图识别-语义槽填充联合模型

² IOB指的是一个词是某一语义槽的开始(begin)、延续(inside)或者非语义槽(outside)。

基于语言理解的工作流程：领域/意图识别、语义槽填充，相关工作默认已经获知了意图识别的结果，并在该意图的语义框架下进行语义槽填充。其中的问题是，如果领域/意图识别错误，会直接影响到语义槽填充任务，从而产生错误传播。为了解决这个问题，很多学者^[27-30]使用了领域/意图识别和语义槽填充的联合模型。其基本思想是使用一个网络结构学习输入的句子信息，同时输出每个词的标签以及句子的领域、意图类别。

引入外部知识

一些工作考虑引入外部知识来提升模型性能。文献[31, 32]等引入了Freebase³或者Wikipedia等外部知识来提升系统的性能。Chen等人^[33]在Encoder-Decoder框架下，引入依存句法分析、语义依存分析等先验知识在数据集较小的情况下能够显著提升模型效果。

数据集

ATIS数据集

任务型对话语言理解的研究工作中，使用最广泛的数据集是航空旅游信息系统(Airline Travel Information System, ATIS)的数据^[34]。该数据集是采集自真实预订飞机票录音，再经过转写后的文本。

ATIS在语言理解任务上使用最广泛的版本⁴是来自于He等人^[10]和Raymond等人^[14]的工作，其中训练数据共4978句，来自ATIS2⁵和ATIS3⁶数据集；测试数据包含893句，来自ATIS3 NOV93和DEC94数据集。其中包含了127种语义槽标签和18种意图。

其他数据集

在一些研究工作中，使用了非公开的数据集。例如Guo等人的工作中，使用了微软Cortana英文对话数据集^[27]。Yao等人使用了微软必应查询(Bing

Query)理解英文数据集^[35]。

在Luo等人的工作中，使用了Nuance公司私有的数据集^[36]。这个数据集包含了200万句训练数据，26万句测试数据，涵盖了170种意图。

对话系统服务平台

如果自己开发一个对话系统，通常需要完成以下工作。

1. 定义和准备：定义该对话系统支持的领域/意图或问题类别，定义语义槽，进行相关的数据标注；

2. 训练和实现：训练领域/意图或问题分类模型，实现语言理解、对话管理和语言生成模块；

3. 部署：在自己的服务器上，部署实现好的工程，对用户提供服务。

在不同的对话系统中，由于领域不同，第一步的定义和数据标注通常会有区别；而第二步和第三步的工作往往是重复的，所以可以考虑将这两步抽象为一个通用工作流程，作为一个算法的黑箱，为开发者提供一个通用的对话系统服务平台。在该平台上，用户只需完成第一步的任务定义和数据标注工作，后面两个步骤由平台自动提供，因此极大地简化了对话系统的开发和部署流程。

表3 对话系统服务平台对比

系统名称	语言理解					对话管理	对话生成
	意图识别	实体/语义槽识别	系统内置实体数量	自定义实体	自定义特征		
LUIS.AI	√	√	中	√	√	×	×
API.AI	√	√	多	√	×	√	√
Wit.ai	√	√	多	√	×	√	√

³ Freebase是一个类似维基百科(Wikipedia)的创作共享类网站，所有内容都由用户添加，采用创意共用许可证，可以自由引用。两者之间最大的不同在于，Freebase中的条目都采用结构化数据的形式。

⁴ <https://www.dropbox.com/s/3lx19jsbw0j7h8a/atls.pk?dl=0>

⁵ <https://catalog.ldc.upenn.edu/LDC93S5>

⁶ <https://catalog.ldc.upenn.edu/LDC94S19>

各大互联网公司纷纷推出了各自的对话系统服务平台,如微软推出的 LUIS.AI,谷歌收购的 API.AI 和脸书收购的 Wit.ai 等。

表 3 是对上述三个平台的主要功能支持度进行的对比。

三个平台都提供了基本的语言理解功能,基本上可使开发者便捷地开发出一个人机对话系统。■



车万翔

CCF高级会员。哈尔滨工业大学副教授、博士生导师。主要研究方向为自然语言处理。
car@ir.hit.edu.cn



张宇

CCF高级会员。哈尔滨工业大学教授。主要研究方向为信息检索、自动问答。
yzhang@ir.hit.edu.cn

参考文献

- [1] Tur G, De Mori R. *Spoken language understanding: Systems for extracting semantic information from speech*[M]. John Wiley & Sons, 2011.
- [2] Mesnil G, Dauphin Y, Yao K, et al. Using recurrent neural networks for slot filling in spoken language understanding[J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 2015, 23(3): 530-539.
- [3] 张志昌, 张宇, 刘挺, 等. 基于线索词识别和训练集扩展的中文问题分类[J]. *高技术通讯*, 2009, 19(2): 111-118.
- [4] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[OL]. (2014). arXiv preprint arXiv:1404.2188.
- [5] Mou L, Peng H, Li G, et al. Discriminative neural sentence modeling by tree-based convolution[OL]. (2015). arXiv preprint arXiv:1504.01106.
- [6] Komninos A, Manandhar S. Dependency based embeddings for sentence classification tasks[C]// *Proceedings of NAACL-HLT*. 2016: 1490-1500.

- [7] Pieraccini R, Tzoukermann E, Gorelov Z, et al. A speech understanding system based on statistical representation of semantics[C]// *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE Press, 1992, 1(1): 193-196.
- [8] Wang Y Y, Deng L, Acero A. Spoken language understanding[J]. *IEEE Signal Processing Magazine*, 2005, 22(5): 16-31.
- [9] Macherey K, Och F J, Ney H. Natural language understanding using statistical machine translation[C]// *Proceedings of the European Conference on Speech Communication & Technology*, 2001: 2205-2208.
- [10] He Y, Young S. A data-driven spoken language understanding system[C]// *IEEE Workshop on Automatic Speech Recognition and Understanding*. IEEE, 2003: 583-588.
- [11] Lafferty J D, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]// *Proceedings of the Eighteenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc. 2001: 282-289.
- [12] Wang Y Y, Acero A. Discriminative models for spoken language understanding[C]// *Proceedings of the International Conference on Spoken Language Processing*. DBLP, 2006: 1766-1769.
- [13] Wang Y Y, Deng L, Acero A. Spoken language understanding[J]. *Signal Processing Magazine IEEE*, 2005, 22(5): 16-31.
- [14] Raymond C, Riccardi G. Generative and discriminative algorithms for spoken language understanding[C]// *Proceedings of the 8th Annual Conference of the International Speech Communication Association*. 2007: 1605-1608.
- [15] Viterbi A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm[J]. *IEEE Transactions on Information Theory*, 1967, 13(2): 260-269.
- [16] Liu J, Cyphers S, Pasupat P, et al. A conversational movie search system based on conditional random fields[C]// *Proceedings of the 13th Annual Conference of the International Speech Communication Association*. 2012: 2454-2457.
- [17] Jeong M, Lee G G. Structures for Spoken Language Understanding: A Two-Step Approach[C]// *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2007: IV-141 - IV-144.

更多参考文献: www.ccf.org.cn/sztsg/cbw/zgjsjxhtx/