

**《软件项目管理》课程项目报告**



题 目  **智能聊天机器人**

学 院  **软件学院**

专 业  **软件工程**

学生姓名  **谭兴成**

学 号 **2018141463120**

年 级 **2018**

二Ο二一 年 六 月

1 绪论

* 1. 项目背景

聊天机器人应用范围广泛，作用明显。预计到2020年，聊天机器人将为85%的客户服务交互提供助力，到2022年，聊天机器人每年将节约80多亿美元的成本。随着聊天机器人技术的不断优化，聊天机器人的应用领域也逐渐广泛，包含金融、电信、旅游甚至体育、医疗等领域，从而为用户打造更好的服务。聊天机器人已开始慢慢渗透到了人们的日常生活中，未来将会有更多细分领域下的应用场景落地，并且在技术、企业和资本的不断助力下，聊天机器人将不断释放其潜力价值。

1.2 国内外研究现状

国外的bot startups种类比较多，各个等级的企业都有，从最上面的应用层来说，slack、messenger、telegram、kik等各个message平台上都有大量的bot，包括各种各样的服务。这类bot门槛较低，缺乏核心技术，通常是一个idea来支撑整个企业，容易同质化，来源可能是各种bot比赛的产物，域名都是.ai，稍微大一点的支持多个平台，很多都是只在slack上使用，有一种bot成海的感觉，什么样的服务都可以用bot来做，强行改变交互方式。有的slack bot服务于team，有的是将slack与其他服务，比如google analytics，以bot的形式进行桥接。目前，我国企业聊天机器人重点布局客服领域。埃森哲与蚂蚁金服发布《新客服白皮书》显示，目前新客服创新中核心技术之一便是对话机器人。报告显示，大量资本将聊天机器人持续投入智能客服领域。截至2018年12月31日，国内共有73家提供数字客服解决方案的企业获得融资，26.03%的企业获得A轮融资，部分公司融资到C轮以上。

1.3 项目的主要工作

我们的项目是用python实现智能聊天机器人，现阶段定下来的主要目标为，实现中文的语音识别，并且能够根据语音完成三种不同情境下的聊天需求和智能助理服务，包括任务完成类型，例如：人：今天天气如何 机器人：今天深圳天气晴朗，25~33度；闲聊类型，人：哇，天气好好，做点什么好呢。机器人：不如出去公园散散步玩玩吧~；问答类型，人：哪个地方？机器人：给你推荐附件的市民中心吧。

可以将任务划分为以下几个方面：

声学语音部分包括：1.语音识别功能(ASR) 2.语音合成(TTS)

自然语音理解部分包括:1.语言理解(NLU) 2.对话状态管理（DST) 3.动作候选排序（Policy） 4.语音生成（NLG）

进阶目标是实现更加智能化的语音对答，包括更多情境下的回复，以及支持更多语言。

1.4 项目组成员及分工

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 技术水平 | 角色 | 工作描述 |
| 蒋祎 |  | 项目管理、前期分析、设计 | 分析系统需求、项目计划、项目团队管理、检查进度 |
| 娄泰宇 |  | 分析、设计 | 分析新功能、软件框架扩展、代码模块分配、设计说明书 |
| 张天翊 |  | 设计、编码 | 数据交换、模块设计、接口设计、模块编写 |
| 谭兴成 |  | 编码、测试 | 模块编写、加载程序编写、安装程序编写、软件测试 |
| 王世龙 |  | 测试、文档编写 | 软件测试、软件测试文档、用户操作手册 |

2 相关技术介绍

本项目语音识别部分通过采用卷积神经网络（CNN）和连接性时序分类（CTC）方法，使用大量中文语音数据集进行训练，将声音转录为中文拼音，并通过语言模型，将拼音序列转换为中文文本。

我们使用的的声学模型采用了深度全卷积神经网络，直接将语谱图作为输入。模型结构上，借鉴了VGG，这种网络模型有着很强的表达能力，可以看到非常长的历史和未来信息，相比RNN在鲁棒性上更出色。在输出端，这种模型可以和CTC方案可以完美结合，以实现整个模型的端到端训练，将声音波形信号直接转录为中文普通话拼音序列。在语言模型上，通过最大熵隐含马尔可夫模型，将拼音序列转换为中文文本。

3 设计工作和进展情况

3.1 项目概况

3.1.1 项目特点

参见项目技术介绍

3.1.2 功能需求

SRS-0010 语音转文字需求（核心功能）

在该系统中将输入的语音转换文字，并能够输出成一定格式和规范的文字数据，提供给下一个功能进行识别和转换。

#

SRS-0020 文字转拼音需求（核心功能）

系统能够将语音转文字中转换出来的文字进行识别，并将其转换为拼音数据，拼音数据将提供给下一个拼音转语音的功能进行识别和转换。

#

SRS-0030 拼音转语音需求（核心功能）

系统能够将语音转文字中转换出来的文字进行识别，并将其转换为拼音数据，拼音数据将提供给下一个拼音转语音的功能进行识别和转换。

SRS-0030 NLP需求（自然语言处理）

让计算机接受用户自然语言形式的输入，并在内部通过人类所定义的算法进行加工、计算等系列操作，以模拟人类对自然语言的理解，并返回用户所期望的结果。

拓展功能：针对不同情境 做出不同的回应

#

SRS-0040 问答类型聊天需求

系统应提供用户的问答型对话需求，例如：XXX，今天天气如何？机器人：今天天气晴，体感温度16摄氏度，建议您多加一件衣服。等等类似的对话。

#

SRS-0050 闲聊类型聊天需求

系统应提供用户的闲聊类型聊天需求，例如：机器人：XXX，您今天看起来心情很好啊，是发生了什么好事情么。等等类似的对话。

#

SRS-0060 情感分析

系统应可以从与用户的对话当中，提取有效信息并分析出用户当前的心情概率，并以此为依据进行闲聊类型聊天的方向选择。例如：当判断用户表现出明显的不耐烦或者其他消极情绪时，要及时停止话题结束对话，等

#

未来发展方向：

SRS-0070 智能家居

系统应在不同问答类型聊天的基础上，进行与家居设备的操作互动。例如，XXX，我想听《xxxx》音乐，当用户提出这样的目的性问答聊天时，系统需要进行回应，并进行音乐播放的操作。类似的还有，开门，关灯等，达到智能家居的效果。

SRS-0080 微信小程序聊天机器人

系统应搭建在微信小程序平台上，能够支持语音输入和键盘输入，并且能够在不同情境下进行反馈。

3.2 项目开发计划及阶段性完成情况

3.2.1 工作内容

产品范围界定：语音识别，语音合成，语言理解，对话状态管理，动作候选排序，语音生成。

工作范围界定：规范的项目文档，可运行的程序（良好的注释）

3.2.2 工作进度安排

第1周-第2周：学习网络上的模型

1. 第5周：模型初步建立
2. 第8周：逐步完善模型

第9-第12周：连接三个模块整合成为最终项目

最终提交：可运行代码和项目文档

3.3 项目的开发内容和结果

3.3.1 系统需求分析

实现中文的语音识别，并且能够根据语音完成三种不同情境下的聊天需求和智能助理服务，包括任务完成类型，例如：人：今天天气如何 机器人：今天深圳天气晴朗，25~33度；闲聊类型，人：哇，天气好好，做点什么好呢。机器人：不如出去公园散散步玩玩吧~；问答类型，人：哪个地方？机器人：给你推荐附件的市民中心吧。

可以将任务划分为以下几个方面：

声学语音部分包括：1.语音识别功能(ASR) 2.语音合成(TTS)

自然语音理解部分包括:1.语言理解(NLU) 2.对话状态管理（DST) 3.动作候选排序（Policy） 4.语音生成（NLG）

进阶目标是实现更加智能化的语音对答，包括更多情境下的回复，以及支持更多语言。

3.3.2 系统设计

体系架构设计

智能聊天机器人系统总共分为三大模块，分别为语音识别模块：识别用户的语音输入并进行拼音处理；AI图灵模块：进行内层逻辑运算，包括语音分析，情感分析等；语音合成模块：将得到的结果通过语音合成的方式输出。

人机界面设计



系统接口设计

内层逻辑使用使用图灵AI接口，外层界面使用QT接口UIform界面设计

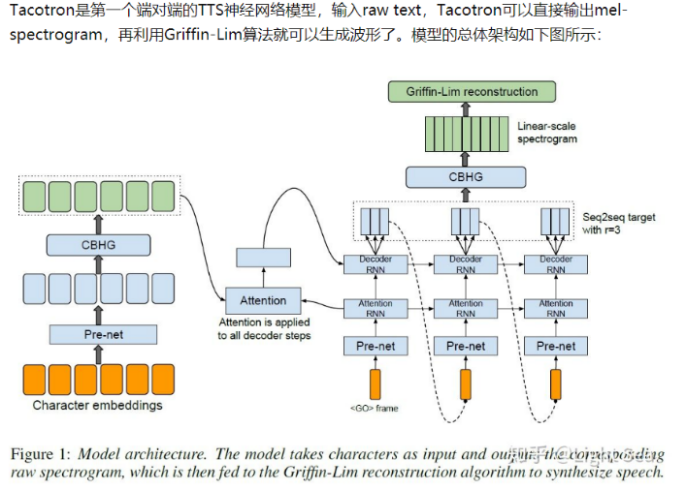
3.3.3 系统具体实现

语音识别部分通过采用卷积神经网络（CNN）和连接性时序分类（CTC）方法，使用大量中文语音数据集进行训练，将声音转录为中文拼音，并通过语言模型，将拼音序列转换为中文文本。

我们使用的的声学模型采用了深度全卷积神经网络，直接将语谱图作为输入。模型结构上，借鉴了VGG，这种网络模型有着很强的表达能力，可以看到非常长的历史和未来信息，相比RNN在鲁棒性上更出色。在输出端，这种模型可以和CTC方案可以完美结合，以实现整个模型的端到端训练，将声音波形信号直接转录为中文普通话拼音序列。在语言模型上，通过最大熵隐含马尔可夫模型，将拼音序列转换为中文文本。

内层逻辑直接调用图灵AI的api，从而实现内部的回答逻辑

采用TACOTRON 模型，TACOTRON是一个端到端的TTS模型，模型核心是seq2seq + attention。模型的输入为一系列文本字向量，输出mel-spectrogram, 然后在使用Griffin\_lim算法生成对应音频。模型结构如下图：



3.3.4 系统测试

采用黑盒测试，通过不同的语音输入，根据最后返回的语音判断测试结果

首先是基本的对于语音合成和语音识别的测试

其次是针对问答内容的测试，例如：人：今天天气如何 机器人：今天深圳天气晴朗，25~33度；闲聊类型，人：哇，天气好好，做点什么好呢。机器人：不如出去公园散散步玩玩吧~；问答类型，人：哪个地方？机器人：给你推荐附件的市民中心吧。

具体参考测试视频

4 讨论与体会

4.1 对项目过程的体会

我们小组的选题是聊天机器人，这个题目对我们来说有很大难度，选择的时候其实也是抱着挑战一下和重在学习过程的心态。确定了选题之后我们小组经历了很长一段时间的迷茫期，毕竟我们几乎没有接触过一个完整的人工智能项目，感到无从下手。于是我们寄希望于在网上找一些现有的项目，学习一下语音识别和语音合成的原理、过程。尽管我有一些深度学习和数据挖掘的基础知识，但是也仅仅局限于一些简易的机器学习算法，学习的东西相对来说比较零碎，面对代码中需要用到的成体系的数据处理流程和模块调用感到力不从心。在学习人工智能相关知识的过程中，对于那些“深入浅出”的讲解我能够理解，但是真正要我用代码实现的时候难度就体现出来了。这也是对我的一个警示，那就是既要关注书本上的知识和题目，也要加强自己的实际动手能力，作为软件工程专业的学生，动手写代码是重中之重。

项目管理的团队合作也十分关键，要在有效沟通的基础上进行分工，同时分工要明确，把任务落实到每个人。越是复杂、难度高的项目越需要良好的小组分共与合作。比如，我们这个项目中分为总的来说可以分为语音识别、语音合成、和图灵AI这三个模块，从一开始，我们小组就将任务根据各人的能力和兴趣分配给了每个人。但是因为没有相关项目经验以及项目难度估计不到位，导致后面需要重新分配工作。同时，尽管有了分工，但小组成员也不是各做各的，大家在学习的过程中经常互相交流经验，探讨对某个算法原理的理解以及对项目的改进想法等，正是因为有了大家的团队协作，我们的项目才能不断推进下去。

4.2 对项目的评价

人工智能是一个比较火热、又稍有难度的题材，当初我们小组之所以将这个作为我们的选题，也是想要挑战一下自己，希望自己能够从这次项目中对人工智能有更深刻的认识以及更多的实践经验。考虑到独立做出一个完整的人工智能难度很大，我们小组的目标并没有很高，能通过找一些现有的模型和数据、了解代码的功能和结构、对各部分功能的代码进行整合、最终实现一个低配版的聊天机器人就行了，最重要的目的还是希望在学习过程中对人工智能有更进一步的认识以及体验项目实践和管理的过程。聊天机器人中的核心语音识别和语音合成模块是比较成熟的技术，有许多现有的项目和代码可供参考，通过对不同模型和学习和比较，我对语音识别和语音合成的相关原理、架构有了更加清晰的认识。最终我们实现了语音合成和语音识别功能，而内层逻辑选择调用图灵AI接口来实现，最后加上了图形化界面，完成了一个简易版的聊天机器人。

5 小结

经历了这次智能聊天机器人项目，我从中感受到最深的一点，就是我的动手能力十分欠缺，以及知识成体系化做得不够好。在学习机器学习相关知识的过程中，我感觉自己理论知识学习还算过得去，但是真正到了写代码实践的时候却发现许多地方都有问题，属于是平时学习中觉得大致了解了就不去深入思考的后果。同时，我平时学习的知识也偏零碎化，学习了各种原理和算法，但是并没有去实践操作一个完整的数据处理流程，导致在做完整项目的时候无从下手。

其次就是通过这一学期的课程，我体验到了完整的项目管理过程，需求分析、小组分工、设计、代码编写、测试、文档编写等。越是复杂和困难的项目，项目前做的准备工作就越多，小组分工和设计也要充分评估工作量和难度后再做出决定，一但编码工作正式启动，再去修改需求和设计，成本就会增加许多了。项目的进度管理也是决定项目完成度的指标，这次项目中，也有未能按照计划在规定时间内完成对应任务的情况，这里面有难度评估不准确的原因，也有自身松懈、拖延的原因，可见严格执行项目计划安排也是很重要的。

参考资料

附录1项目开发计划

1 引言

2 项目概述

3 实施计划

4 支持条件

5 专题计划要点

附录2需求规格说明书（包括需求分析模型-功能、数据、行为模型等）

1 引言

2 任务概述

3 需求规定

4 运行环境规定

附录3设计文档（包括设计模型-软件架构、数据结构、用户界面、功能模块等）

概要设计说明书

1 引言

2 总体设计

3 接口设计

4 运行设计

5 系统出错处理设计

详细设计说明书

1 引言

2 系统的结构

3模块1设计说明

4模块2设计说明

5模块3设计说明

。。。

附录4用户文档说明书（包括软件的描述、安装和使用说明等）

1 引言

2 软硬件环境

3 安装说明

4 操作说明

5 功能列表