

**《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  **娄泰宇**

学 号 **2018141463052**

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

第6-第8周：逐步完善模型

1. 第12周：连接三个模块整合成为最终项目

第12-第13周：测试，进行最终完善

最终提交：可运行代码和项目文档

小组成员具体分工：

第1周-第2周：蒋祎，张天翊学习语音合成模型，确立项目语音合成所用模型；娄泰宇，谭兴成，王世龙学习语音识别模型，确立项目语音识别所用模型

第3-第5周：语音合成模型学习完毕，确立为Tactorn，语音识别模型学习完毕，确立为Aspeech；进行模仿学习和代码重构，形成自我项目并进行debug。蒋祎，张天翊负责语音合成，娄泰宇，谭兴成，王世龙负责语音识别

第6周-第8周：进行测试训练，根据效果逐步修改和完善模型具体细节，根据情况选择是否需要更换模型。蒋祎开始负责项目整合工作，学习图灵AI接口。

第9周-第12周：蒋祎连接三个模块整合成为最终项目

第13-第43周：蒋祎，谭兴成，王世龙负责测试，进行最终完善

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

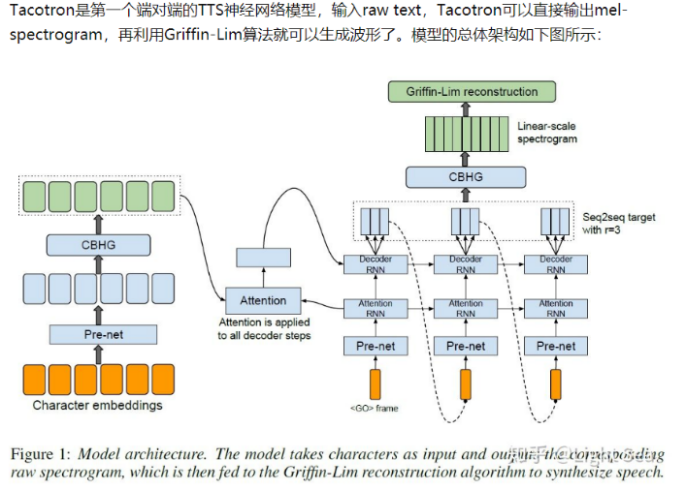
3.3.3 系统具体实现

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

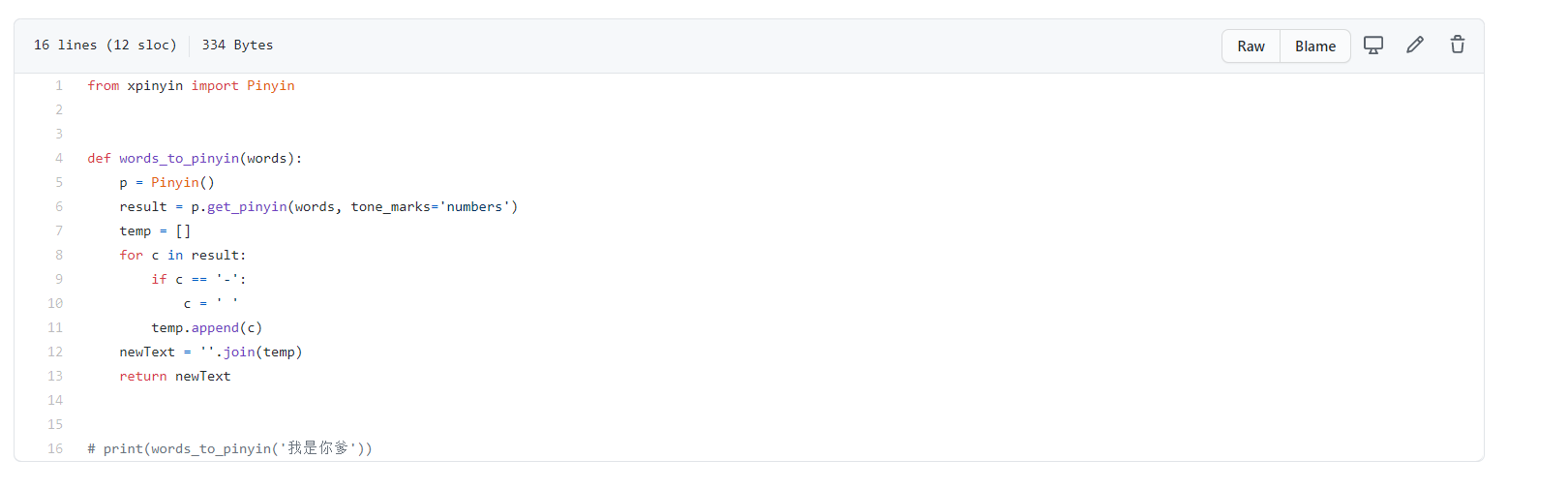
采用TACOTRON 模型，TACOTRON是一个端到端的TTS模型，模型核心是seq2seq + attention。模型的输入为一系列文本字向量，输出mel-spectrogram, 然后在使用Griffin\_lim算法生成对应音频。模型结构如下图：



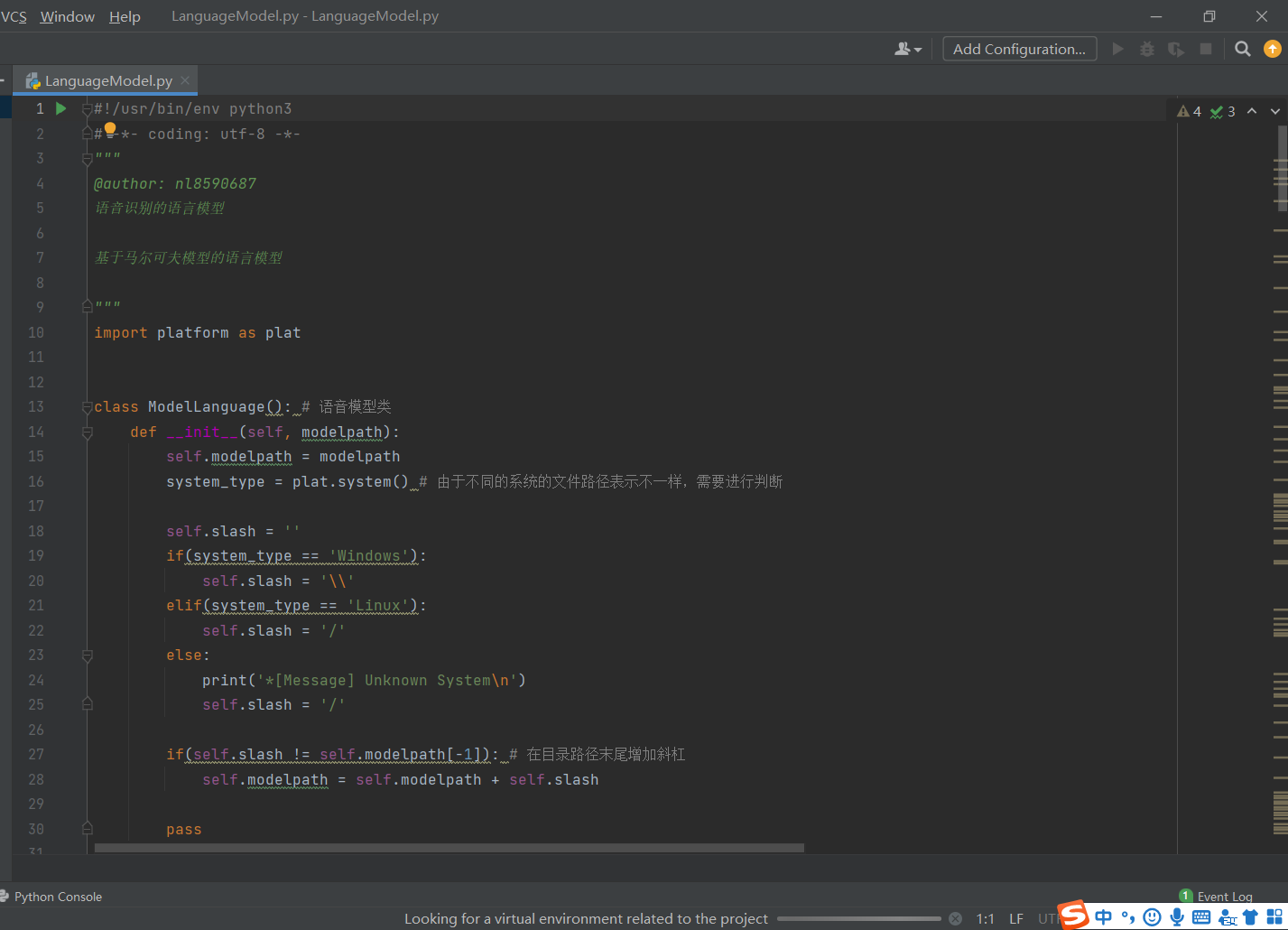
语音识别模型代码 eval.py文件：



转接口，文字转拼音部分桥接代码 文字转拼音.py文件：



语音识别模型代码 LanguageModel.py文件



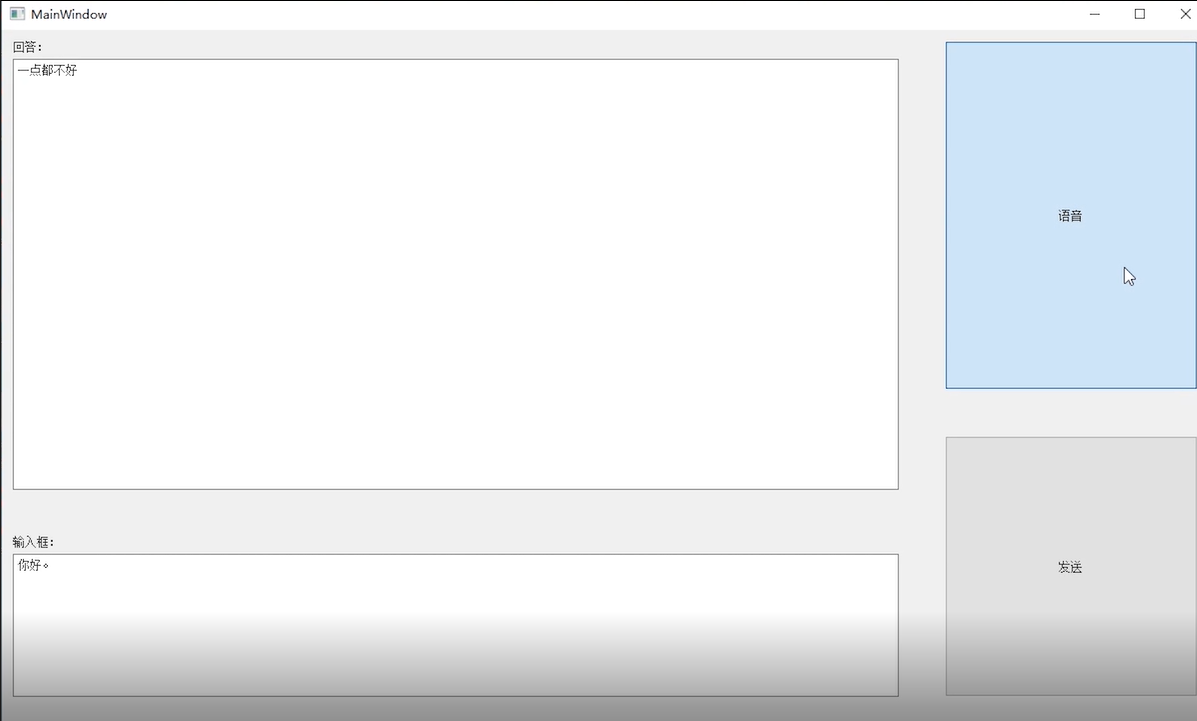
3.3.4 系统测试

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

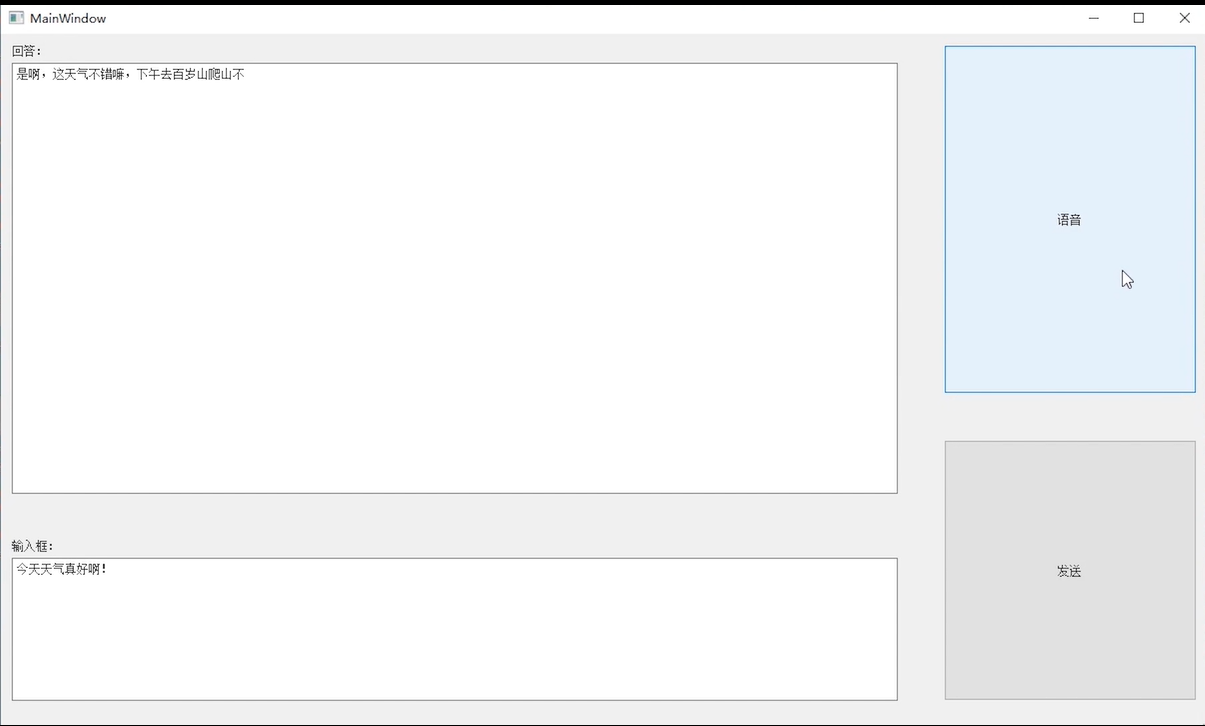
首先是基本的对于语音合成和语音识别的测试

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。

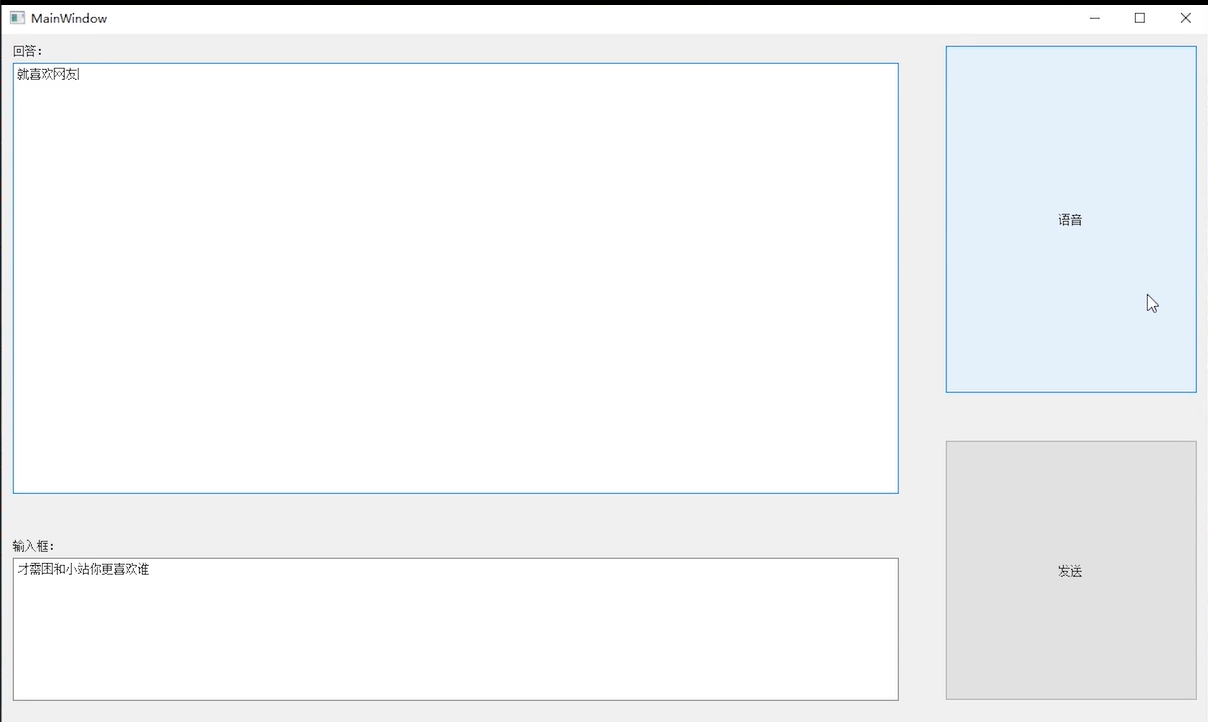
闲聊类型提问：你好；语音输入正确。回答：一点都不好，逻辑稍有问题



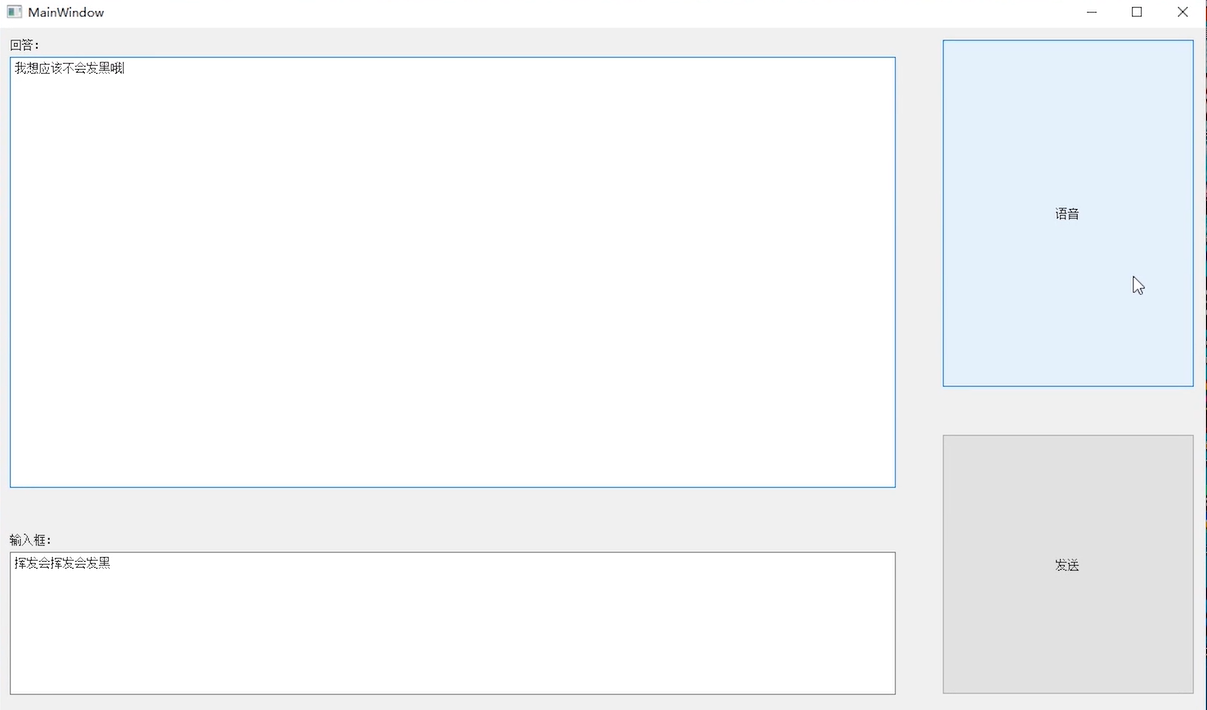
闲聊类型提问：今天天气真好；语音输入正确。回答：天气不错，逻辑正确



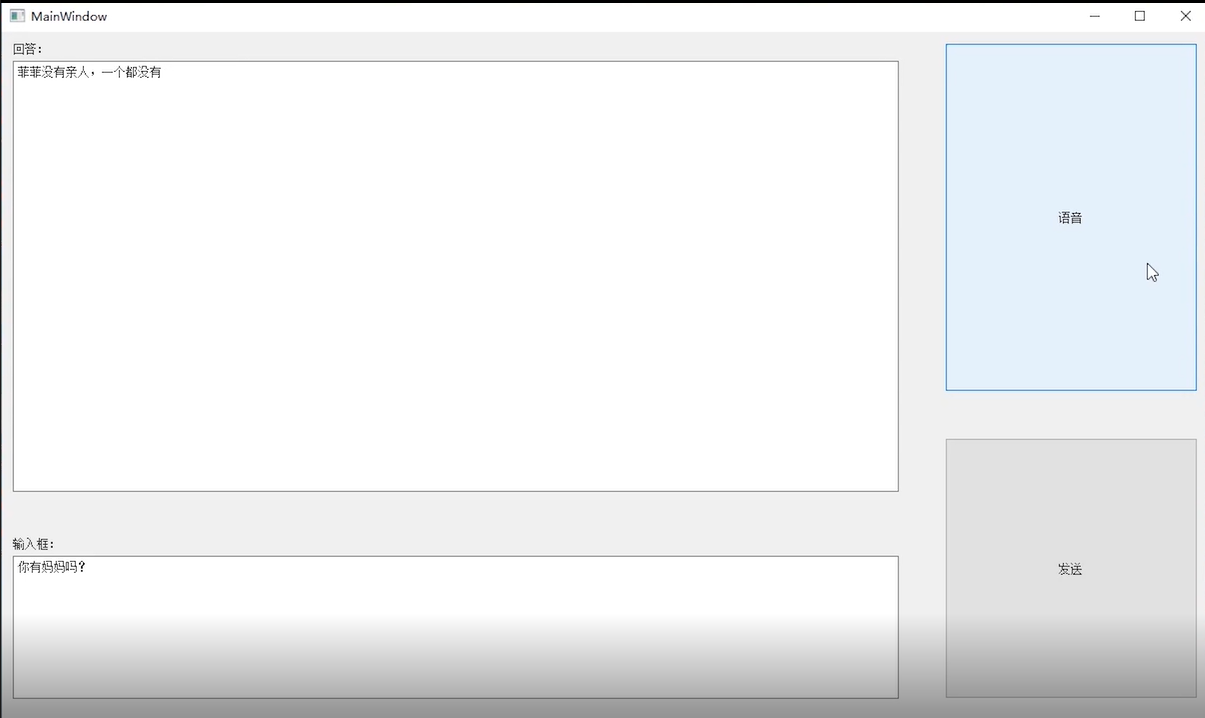
提问类型：蔡徐坤和肖战你更喜欢谁；语音输入有误，人名未能准确翻译；回答：就喜欢网友，逻辑有误，猜测是读出了人名但没有读出具体信息。



绕口令：黑化肥挥发会发黑，测试语音输入；语音输入有误，未能准确断句和翻译句子；回答：不会发黑，逻辑尚可，猜测是读出了发黑并以此为问题所回答。



提问类型：你有妈妈么，语音输入正确；回答：菲菲没有亲人，逻辑正确



具体可见测试视频

4 讨论与体会

4.1 对项目过程的体会

这次做的项目是智能聊天机器人，我和另外一个组员一起分配到了语音识别的模块，这个领域我们都是第一次接触，刚刚开始的时候做的磕磕绊绊的，查阅了大量的资料也一筹莫展，所以最后还是决定找到现有的代码，借鉴和学习。虽然这么说，但是做起来也还是很慢，很困难，并且没有适合的电脑去运行和跑模型，学校里面每天晚上都要断电，自己跑出来的只有几个epoch的完全不够用，准确率和识别率低的离谱，只有在语音转拼音的部分能够达到70%-80%左右的准确率，但是拼音转文字就会低的可怕。等回过神来，已经没有什么时间去对接了，然后老师提醒我们将识别和转换等的功能分开，只要能够把核心的功能实现了就行，当然尽量应该在过程中自己一步步去做，去学习。

做到最后，我能够感受到这是一个很广泛的领域，加了一些QQ交流群，提出的问题里面的人也都很热心地给我解答，现有的市场还没到饱和的程度，不过智能聊天机器人能够应用到的领域基本上也被摸得七七八八的了，再次发展可能期待更新的技术带来的一波革命。

这门课程让我学习和收获到了很多，老师和助教定期检查项目成果，督促和提出一些宝贵的修改意见，老师也很耐心地对待我们，即使是没有完全成功的项目，老师也鼓励和督促我们，看重过程的课程教学让人感觉很舒适，学的很有动力，感谢老师的辛勤付出，也感谢助教老师。

4.2 对项目的评价

项目的选题是有挑战性的，毕竟是我们都没尝试过的领域，但也是可控的，因为网上有很多的教程可供学习，甚至有现成的代码和模型，从项目中学到了很多也体会到了很多，毕竟结对编程不同于个人编程，瀑布式的开发最后很可能每个人写着写着写出了很多个不同理解的版本，所以在一开始明确需求并且确定每个人该干什么，该完成什么，项目的范围和边界在哪是很重要的一个环节。

感谢老师的指导和教学，这门课的开设应该也是为了让我们熟悉和了解软件项目的开发的整个的一套流程，学习和收获到了很多，感谢老师和助教的辛勤付出！

5 小结

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