

一个基于多线程的聊天机器人

曹胜操

叶天

孙之清

北京大学信息科学技术学院

一个基于多线程的聊天机器人

我们将创建一个基于多线程的聊天机器人。通过巧妙的线程调度，我们的机器人能够做到以下几点普通的聊天机器人所不能做到的功能：

- 1、能够在自己说话时被用户的语音打断
- 2、能够分辨出自己的语音中夹杂着的用户语音的意思
- 3、能够自主的判断什么时候用户开始和结束说话并给出回应

线程设计

线程设计是该聊天机器人的主要创新之处。我们选择使用不同的线程来实现机器人的各种功能。线程的创建、调度和管理可以通过用户级线程 API 有效地实现，并且线程之间在数据的传递上比进程间通信要方便很多，这是选择使用线程来实现的重要原因。

线程的种类及执行的操作

在这一部分，我们列举出所设计的不同线程的种类，以及它们分别要执行的操作。我们要实现的聊天机器人主要有以下 5 种线程：

Main 线程.

主线程，在系统启动后初始化整个系统，并负责各个线程的创建、调度和回收

Listen 线程.

收集并分析外界的声音，通过滤波等有关信号的处理，得出用户的声音音频，并将数据发送给 Analyze 线程

Analyze 线程.

a. 将从 Listen 线程那里收集到的用户音频数据发送给科大讯飞，得到音频所对应的文字

b. 将 a 步骤得到的文字发送给图灵机器人，得到根据所发送的文字而产生的回答文字

c. 将 b 步骤得到的回答文字发送给科大讯飞，得到回答文字所对应的音频

d. 将 c 步骤得到的回答的音频数据传送给 **Speak** 线程和 **GUI** 线程

Speak 线程.

根据从 **Analyze** 线程那里得到的机器人回答的音频，执行播放音频的操作

GUI 线程.

根据从 **Analyze** 线程那里得到的机器人回答的音频，控制屏幕上的虚拟人物形象，使得虚拟人物形象产生与回答对应的口型和动作

线程工作状态

介绍完不同种类的线程以及相应执行的操作之后，我们需要设计机器人系统的状态及其转移。在每个状态中，将会有不同的线程组合进行工作。

Init 状态.

a. **Main** 线程初始化整个系统

b. **Init** 状态工作完成后，进入 **Listening** 状态

Listening 状态.

a. 创建并运行 **Listen** 线程

b. **Listen** 线程收集并分析外界的声音，处理出用户的声音后，将数据发送给

Analyze 线程

c. 进入 Listening + Analyzing 状态

Listening + Analyzing 状态.

a. Listen 线程持续收听外界的声音

b. 创建并运行 Analyze 线程，与云端进行多次交互，从上一状态中的用户音频最终得到机器人对应的回答音频

c. 在进行分析时需要保存和利用相关历史信息，提高上下文逻辑性；这一点在云端已经有所实现

d. Analyze 线程分析完毕后，将机器人回答的音频交给 Speak 线程

e. 进入 Listening + Speaking 状态

Listening + Speaking 状态.

创建并运行 Speak 线程和 GUI 线程，机器人开始播放回答的音频，并同时在屏幕上显示虚拟人物形象

b. Listen 线程继续收听用户声音，此时由于机器人本身也在说话，因此需要通过信号处理的方式过滤机器人自身的声音，保证能够得到用户的声音

c1. 若 Speak 线程顺利完成，则进入 Listening 状态

c2. 若在 Speak 线程运行的过程中，发现用户正在说话，则结束 Speak 线程和 GUI 线程，保留本次要说的内容，进入 Listening + Analyzing 状态

Finish 状态.

a. 在上述任意状态中，如果用户关闭聊天程序，则进入 Finish 状态

b. 将正在运行的各线程回收

功能实现

该聊天机器人的各种功能的实现主要依赖于各种不同厂家提供的 API 服务，下面将从各个方面阐释我们功能的实现方式。

环境

我们将在 Windows 平台上实现我们的聊天机器人。由于我们需要实现语音形式的聊天，聊天机器人必须要能够使用麦克风和扬声器来分别输入和输出语音。对于语音和文字信息的处理需要用到在线的 API，所以聊天机器人还需要能够实时联网。除此之外，我们还将渲染出虚拟的表情图像，聊天机器人需要具有图像渲染的能力。

语音

讯飞开放平台是一个以语音交互为核心的人工智能开放平台，它提供了优秀的语音听写、语音合成技术，可以为我们所用。讯飞开放平台提供了一套完善的 SDK，利用其中的接口我们就可以编写自己的程序，实现语音和文本之间的转换。在聊天机器人运行的过程中，完成了对音频的基本处理之后，我们就可以把音频发送给平台，从平台接收语音听写得到的文本。经过其它的处理之后得到应答的文本，我们再使用另一套类似的方式将文本发送给平台，得到语音合成的音频。

文本

图灵机器人是一个基于自然语言处理等技术的平台，可以提供具有流畅自然的中文对话回答功能的机器人，有着准确分析语义、学习上下文场景的能力。图灵机器人提供了一套简洁易用的网络 API，通过发送 json 请求就可以得到机器人的回复。在聊天机器人运行的过程中，得到用户所说的文本之后，我们就可以通过网络请求利用该平台得到一个期望的回答文本，然后再进行后续的处理即可。该平台提供的机器人可以进行有上下文环境的聊天，这样就可以根据用户之前所说的历史信息来进行更准确的应答。此外，机器人还可以具有其它一些扩展功能，如数字计算、天气查询等等，给用户提供一些更丰富的体验。

图像

MikuMikuDance 是一款轻量级的 3D 动画制作软件，在导入已经制作好的 3D 人物模型之后，就可以利用 MMD 制作出简单的 3D 动画。我们可以预先渲染好一些虚拟人物形象的基本动作形态，在聊天机器人运行的过程中，根据实时计算得到的文本和语音来动态展现虚拟人物形象，呈现出更加真实有趣的聊天效果。

个性化聊天功能

通过 API 的设定，该聊天机器人可以在聊天的同时提供不同的个性化功能，从而方便人们的生活。

生活百科

提供对生活小常识、生活小百科的数据查询，

数字计算

此功能可进行加减乘除、乘方、开方、指数、对数、统计等多种运算

问答百科

提供对百度百科现有的知识信息的查询，百度百科已经收录了超过 1300 多万的词条，参与词条编辑的网友超过 580 万人，几乎涵盖了所有已知的知识领域。

中英互译

中英互译是一项在线翻译服务，提供中文和英语的互译服务。可以通过使用中英互译功能为用户提供实时优质的翻译服务，提升产品体验。您只需要通过开启此功能，传入待翻译的内容，并指定要翻译的源语言和目标语言种类，就可以得到相应的翻译结果。

脑筋急转弯

此功能随机提供脑筋急转弯及答案。脑筋急转弯最早起源于古代印度。就是指当思维遇到特殊的阻碍时，要很快的离开习惯的思路，从别的方面来思考问题。现在泛指一些不能用通常的思路来回答的智力问答题。

天气查询

可查询全国 2567 个城市、全国 2056 个热门旅游景点和全球天气情况，包括 3 天天气预报、全国灾害预警、实况天气、每小时天气、PM2.5 空气质量、生活指数等丰富数据。

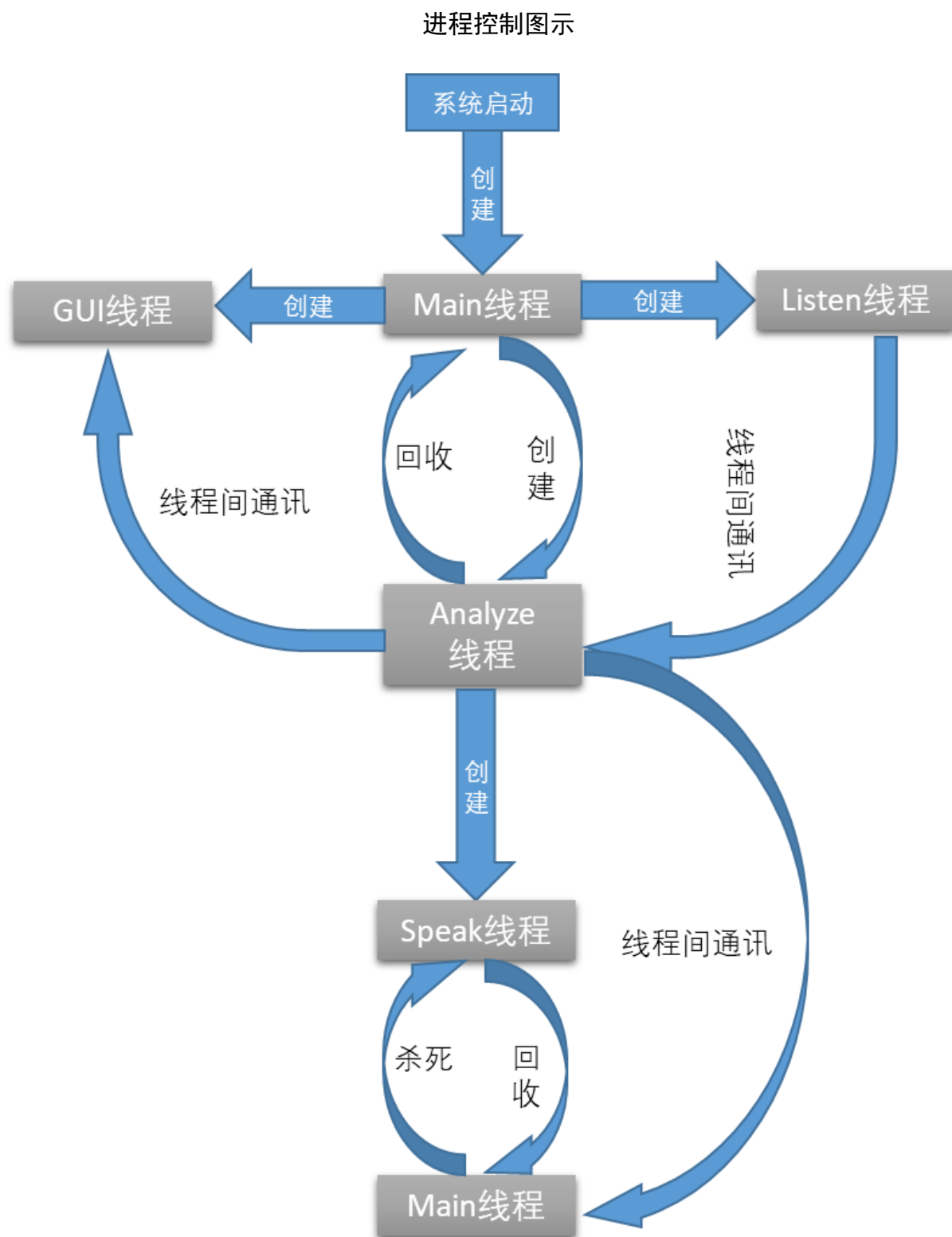


Figure 1. 进程创建与调度