专利技术交底书样本

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 客户名称 |  | | | |
| 发明名称 | 一种基于Multi-Agent为核心的通用语音交互助手 | | | |
| 技术联系人 | 郭鸿宏 | 邮 箱 | weiwang@szlanyou.com | |
| 手 机 |  | | 专利类型 | 发 明  实用新型 |
| 固定电话 |  | |

# 一、技术背景

随着智能设备和人工智能技术的日益普及，语音交互助手逐渐成为我们日常生活中不可或缺的辅助工具。这些助手不仅能够准确识别和响应用户需求，还提供了诸如智能家居控制、音乐播放、知识问答、实时天气查询和新闻资讯等多样化服务，大大提升了生活的便捷性与舒适度。它们丰富了用户的信息获取方式，增强了日常生活的整体体验。

然而，目前市场上的语音助手普遍依赖单一的智能处理单元来执行任务和解析指令，这种设计存在明显的局限性。首先，在处理复杂或多层次的对话时，这些系统通常只能进行简单的规则匹配和命令识别，难以深入解析用户的深层意图。尤其是在同时处理多项任务时，现有系统的表现往往不尽如人意。此外，传统基于规则的语音助手在理解语义和用户情感方面表现不足，使得用户常常感到自己只是在与机械程序互动，而缺乏必要的拟人化体验。

随着语音助手应用环境的日益复杂，单一处理单元的设计已难以满足跨场景应用的需求。因此，发明一种能够实时、动态地处理复杂语境和多任务的语音交互系统变得尤为重要。这将不仅提升用户的交互体验，还能拓展语音助手的应用场景，使其更加智能和人性化。

# 二、发明目的

本发明旨在开发一种基于Multi-Agent系统的语音交互助手，提供高度灵活和可移植的多智能体架构。系统通过整合多个专责的智能处理代理（Agent），并利用优化的意图分类算法，大幅提升了在多场景、多设备复杂应用中的任务分发准确性。多智能体设计不仅有效降低了意图识别模块的负担，还增强了系统的智能表现和用户意图的理解精度。通过高效的协作机 制，系统实现了更加快速和精确的语音理解及响应。此外，结合RAG知识库检索、多引擎爬虫搜索和工具调用模块，系统充分利用丰富的历史数据和实时信息，加强了交互的相关性与准确性。本系统基于强大且灵活的多智能体模型平台构建，具 备出色的适应性和扩展性，适合多种硬件和软件环境。尽管增加了智能体数量，但合理的架构设计避免了链路复杂性的显 著增加，确保用户拥有高效自然的交互体验。系统能够根据具体应用需求快速重组和调整，使产品具有高度的可移植性和 市场适应性，广泛适用于车载助手、智能家居控制、个人生活助理等智能交互场景，满足用户多样化的需求。

# 三、技术方案

本发明所要解决的技术问题是通过以下技术方案实现的，一种基于Multi-Agent为核心的通用语音交互助手，包括以下步骤：

步骤1：系统载体持续循环监听环境，并将监听到的内容传送至唤醒识别模块进行解析。一旦识别到用户设定的唤醒词，系统将发出语音交互提示，引导用户输入，并启动 Multi-Agent 核心。

步骤2：采集用户的语音输入，并根据预设的空白阈值对输入中的空白部分进行过滤。随后，调用 STT（Speech-to-Text，语音转文本）模块将用户的语音输入转换为文本。

其中为采集的用户单轮语音输入，为设定的空白阈值。根据该阈值对输入中的停顿等空白部分进行过滤。例如，若空白阈值为 200ms，则仅过滤 200ms 以上的语音空白。最终，将过滤后的结果转换为文本 text。

步骤3：将前述文本输入存入历史对话数据库，并与历史对话数据一起提交给意图识别模块进行意图分类。根据分类结果，系统将决定是将任务分配给其他代理、调用工具获取扩展信息，还是进入问答流程。具体分类流程如下：

1. 文本输入重构为多轮文本输入。

|  |  |
| --- | --- |
|  | （1） |

其中，为用户语音输入转换后的文本输入， 表示将用户输入存入历史对话数据库，并提取对应用户与会话相关的历史数据。其结果即为历史对话 加上本轮文本输入 。

1. 调用意图识别模块对用户输入进行分发，分发给其它Agent或模块。

具体的计算方式如下：

1）调用微调后的意图分类大模型对重构后的多轮文本输入进行分析，计算每个场景的可能性。

|  |  |
| --- | --- |
|  | （2） |

其中即表示不同场景的可能性向量，场景包括智能家居（Smart Home）、设备操纵（Device Control）、工具调用（Tools）以及问答（QA）。每个维度的值表示该场景的可能性排名，例如，如果本轮输入属于智能家居的可能性排名第二则。

2）针对每个类别的可能用法构造数据，转化为向量并存储在向量库中。再将用户文本输入也转化为向量，并与每个类别的预设向量库计算相似度，得出一个可能性排名即，这个过程即。

|  |  |
| --- | --- |
|  | （3） |

3）对大模型分类结果和预设向量库相似度计算结果进行加权平均得到最终分类结果**。**

|  |  |
| --- | --- |
|  | （4） |

其中，权重的计算权重和的计算权重可以根据经验预设，也可以根据分类表现进行动态调整。

Ⅰ.若最终分类结果为问答QA，即,则跳转步骤4。

表示场景问答QA的加权排名值小于其它所有场景，即场景问答（QA）的可能性最高。下列公式相近，不再赘述。

Ⅱ.若最终分类结果为工具调用（Tools），即,则跳转步骤5。

Ⅲ.若最终分类结果为智能家居（Smart Home），即,则跳转步骤6。

Ⅳ.若最终分类结果为设备操纵（Device Control），即,则跳转步骤7。

步骤4：根据意图识别模块的反馈结果，选择适当的知识补充方式，将重构后的多轮文本输入与补充后的知识一并转交给步骤8的问答大模型进行生成输出。

意图识别模块中的大模型在计算类别时，会提前判断选取何种知识补充方式，并通过 标记位指示是否使用 RAG 方式。如果最终分类场景为问答（QA）且，则启用 RAG 方式进行知识补充。否则，将通过多引擎爬虫搜索补充知识。此外，当 RAG 方式无法召回时，也会切换到多引擎爬虫搜索方式。

知识补充结束后会输出一个多轮文本输入与补充知识Knowledge的集合：

，Knowledge）

步骤5：根据意图识别模块的反馈结果，调用指定工具获取补充信息，并将这些信息与多轮文本输入打包成一个集合，交给步骤8的问答大模型进行生成输出。

|  |  |
| --- | --- |
|  | （1） |
| ，） | （2） |

在意图识别模块中，大模型在计算类别时会提前生成一条工具调用语句。在 Tools 模块中，这条语句会被解析，即通过 进行工具调用。Tools 模块负责解析工具调用语句指定的工具和传递的参数，工具调用结束后返回一个补充信息 。最终生成一个包含多轮文本输入 与补充信息的集合 。

步骤6：智能家居模块会判断意图识别模块的任务派发是否合理，将不合理的派发驳回，对合理的任务派发进行进一步分析，并调用指定家居设备。

智能家居模块本身即为一个以大模型为核心的 Agent 智能体，因此不依赖意图识别模块的分析，可以自行对多轮文本输入 进行分析，判断其是否属于智能家居范畴。

如果属于智能家居范畴，则解析用户需要调用的家居设备及执行的功能，并将用户需求转化为智能家居调用语句并执行，其流程与 Tools 模块类似，封装为。若不属于智能家居范畴，则进行驳回。

驳回有两种方式：对于时间敏感型场景，可以直接驳回给步骤8的问答大模型，以便快速回复用户；对于时间相对不敏感的场景，则驳回给意图识别模块重新定向。系统最多只能重定向一次，若重定向失败则驳回给问答大模型。

步骤7：设备操纵模块会判断意图识别模块的任务派发是否合理，将不合理的派发驳回，对合理的任务派发进行进一步分析，并操纵设备完成指定操作。

操纵模块也是以大模型为核心的 Agent 智能体，因此不依赖意图识别模块的分析，可以自行对多轮文本输入进行分析，，判断其是否属于设备操纵范畴。

如果属于设备操纵范畴，则解析用户需要调用的设备（如手机、电脑）及执行的功能，并将用户需求转化为设备调用语句并执行，流程与 Tools 模块类似，封装为。驳回机制与智能家居模块相同。

步骤8：将上述各步增补好知识的输入或者被驳回的多轮文本输入输入到问答大模型中，流式生成文本输出，并实时交付给 TTS（Text to Speech，文本转语音）模块生成语音回复。

本发明的技术方案还包括一种基于Multi-Agent为核心的通用语音交互助手，其中包括9个主要功能模块，4个模型，2个存储库：

1. 唤醒识别模块：用于监听环境，在触发唤醒词时启动 Multi-Agent 核心。
2. 意图识别模块/意图识别大模型：用于分析用户的文本输入，并将其分类为不同的意图类别。该模块利用一个经过微调的大语言模型（LLM）即意图识别大模型，结合预设的向量库，通过加权平均的方式输出最终的意图分类结果。
3. 知识补充模块：用于根据意图识别结果，选择并获取补充知识。该模块包括 RAG（检索增强生成）和多引擎爬虫搜索两种方式，以确保高效的知识扩展。
4. 工具调用模块：用于解析并执行意图识别模块中生成的工具调用语句，完成特定工具的调用并获取相应的补充信息。
5. 智能家居模块/智能家居大模型：用于判断意图识别模块分发的任务是否属于智能家居控制范畴，并执行相应的家居设备操作。
6. 设备操纵模块/设备操纵大模型：用于分析用户的指令并操作指定的设备，如手机或电脑，完成特定功能的调用。
7. 问答大模型：用于接收并处理所有的多轮文本输入和补充信息，生成最终的文本输出，并实时交付给 TTS 模块进行语音生成。
8. 语音转文本模块：用于将用户的语音输入转换为文本，以便后续处理。
9. 文本转语音模块：用于将系统生成的文本输出转换为语音回复用户。
10. RAG知识库：存储用于检索增强生成的知识数据，支持系统的知识扩展功能。
11. 历史对话数据库存：存储用户的历史对话记录，用于支持多轮对话和上下文理解。

# 四、实施示例

为了进一步描述本发明的技术特点和效果，以下结合附图和具体实施方式对本发明做进一步描述。

图1是本发明一种基于Multi-Agent为核心的通用语音交互助手流程图，如图1所示，该方法包括如下步骤：

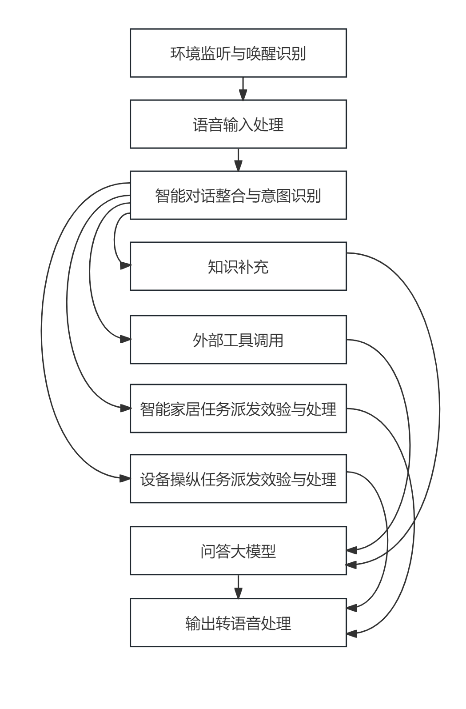


图1

步骤1：系统载体持续循环监听环境，并将监听到的内容传送至唤醒识别模块进行解析。一旦识别到用户设定的唤醒词，系统将发出语音交互提示，引导用户输入，并启动 Multi-Agent 核心。

步骤2：采集用户的语音输入，并根据预设的空白阈值对输入中的空白部分进行过滤。随后，调用 STT（Speech-to-Text，语音转文本）模块将用户的语音输入转换为文本。

其中为采集的用户单轮语音输入，为设定的空白阈值。根据该阈值对输入中的停顿等空白部分进行过滤。

在本实施例中，用户的单轮语音输入 长度为 5 秒，设定的 为 200 毫秒。根据该阈值，过滤用户输入语音中 200毫秒及以上的空白部分。经过识别，共检测到 3 段长度在 200毫秒以上的语音空白，分别为 300毫秒、250毫秒和 450 毫秒。经过空白过滤后，用户输入语音的有效长度缩短为 秒。使用 STT 模块对过滤后的语音进行转换，最终得到用户输入的文本为：

“昨天下了一天的雨真烦，今天天气还不错，帮我打开窗帘透透风。”

即。

步骤3：将前述文本输入存入历史对话数据库，并与历史对话数据一起提交给意图识别模块进行意图分类。根据分类结果，系统将决定是将任务分配给其他代理、调用工具获取扩展信息，还是进入问答流程。具体分类流程如下：

1. 文本输入重构为多轮文本输入。

|  |  |
| --- | --- |
|  | （1） |

其中，为用户语音输入转换后的文本输入， 表示将用户输入存入历史对话数据库，并提取对应用户与会话相关的历史数据。其结果即为历史对话 加上本轮文本输入 。

在本实施例中，用户为第一轮对话，尚不存在历史对话，即，故

1. 调用意图识别模块对用户输入进行分发，分发给其它Agent或模块。

具体的计算方式如下：

1）调用微调后的意图分类大模型对重构后的多轮文本输入进行分析，计算每个场景的可能性。

|  |  |
| --- | --- |
|  | （2） |

其中即表示不同场景的可能性向量，场景包括智能家居（Smart Home）、设备操纵（Device Control）、工具调用（Tools）以及问答（QA）。每个维度的值表示该场景的可能性排名，在本实施例中 3 ]，即大模型判定后续任务最可能需要交付智能家居模块去处理。

2）针对每个类别的可能用法构造数据，转化为向量并存储在向量库中。再将用户文本输入也转化为向量，并与每个类别的预设向量库计算相似度，得出一个可能性排名即，这个过程即。

|  |  |
| --- | --- |
|  | （3） |

在本实施例中， 2 ]，即预设向量库算法也判定后续任务最可能需要交付智能家居模块去处理。

3）对大模型分类结果和预设向量库相似度计算结果进行加权平均得到最终分类结果**。**

|  |  |
| --- | --- |
|  | （4） |

其中，权重的计算权重和的计算权重可以根据经验预设，也可以根据分类表现进行动态调整。

在本实施例中的和基于经验预设为0.8和0.2，=[ 1 , 4 , 2.2 , 2.8 ]

Ⅰ.若最终分类结果为问答QA，即,则跳转步骤4。

表示场景问答QA的加权排名值小于其它所有场景，即场景问答（QA）的可能性最高。下列公式相近，不再赘述。

Ⅱ.若最终分类结果为工具调用（Tools），即,则跳转步骤5。

Ⅲ.若最终分类结果为智能家居（Smart Home），即,则跳转步骤6。

Ⅳ.若最终分类结果为设备操纵（Device Control），即,则跳转步骤7。

在本实施例中，，即经过加权计算融合后，意图识别模块仍然判定后续任务最可能需要交付智能家居模块去处理，跳转步骤6。

步骤4：根据意图识别模块的反馈结果，选择适当的知识补充方式，将重构后的多轮文本输入与补充后的知识一并转交给步骤8的问答大模型进行生成输出。

意图识别模块中的大模型在计算类别时，会提前判断选取何种知识补充方式，并通过 标记位指示是否使用 RAG 方式。如果最终分类场景为问答（QA）且，则启用 RAG 方式进行知识补充。否则，将通过多引擎爬虫搜索补充知识。此外，当 RAG 方式无法召回时，也会切换到多引擎爬虫搜索方式。

知识补充结束后会输出一个多轮文本输入与补充知识Knowledge的集合：

，Knowledge）

在本实施例中，由于，不符合知识补充模块启用条件，因此本轮知识补充模块休眠。

步骤5：根据意图识别模块的反馈结果，调用指定工具获取补充信息，并将这些信息与多轮文本输入打包成一个集合，交给步骤8的问答大模型进行生成输出。

|  |  |
| --- | --- |
|  | （1） |
| ，） | （2） |

在意图识别模块中，大模型在计算类别时会提前生成一条工具调用语句。在 Tools 模块中，这条语句会被解析，即通过 进行工具调用。Tools 模块负责解析工具调用语句指定的工具和传递的参数，工具调用结束后返回一个补充信息 。最终生成一个包含多轮文本输入 与补充信息的集合 。

在本实施例中，由于，不符合工具调用模块启用条件，因此本轮工具调用模块休眠。

步骤6：智能家居模块会判断意图识别模块的任务派发是否合理，将不合理的派发驳回，对合理的任务派发进行进一步分析，并调用指定家居设备。

智能家居模块本身即为一个以大模型为核心的 Agent 智能体，因此不依赖意图识别模块的分析，可以自行对多轮文本输入 进行分析，判断其是否属于智能家居范畴。

如果属于智能家居范畴，则解析用户需要调用的家居设备及执行的功能，并将用户需求转化为智能家居调用语句并执行，其流程与 Tools 模块类似，封装为。若不属于智能家居范畴，则进行驳回。

驳回有两种方式：对于时间敏感型场景，可以直接驳回给步骤8的问答大模型，以便快速回复用户；对于时间相对不敏感的场景，则驳回给意图识别模块重新定向。系统最多只能重定向一次，若重定向失败则驳回给问答大模型。

在本实施例中，由于，智能家居模块启用。经过智能家居模块的大模型校验，当前用户意图属于智能家居模块处理范畴，大模型将分析多轮文本输入 ，结构化的反馈用户需要操作的智能家居以及需要执行的操作：

随后会有脚本分析上述大模型结构化反馈，对指定家居执行对应操作。

步骤7：设备操纵模块会判断意图识别模块的任务派发是否合理，将不合理的派发驳回，对合理的任务派发进行进一步分析，并操纵设备完成指定操作。

操纵模块也是以大模型为核心的 Agent 智能体，因此不依赖意图识别模块的分析，可以自行对多轮文本输入进行分析，，判断其是否属于设备操纵范畴。

如果属于设备操纵范畴，则解析用户需要调用的设备（如手机、电脑）及执行的功能，并将用户需求转化为设备调用语句并执行，流程与 Tools 模块类似，封装为。驳回机制与智能家居模块相同。

在本实施例中，由于，不符合设备操纵模块启用条件，因此本轮设备操纵模块休眠。

步骤8：将上述各步增补好知识的输入或者被驳回的多轮文本输入输入到问答大模型中，流式生成文本输出，并实时交付给 TTS（Text to Speech，文本转语音）模块生成语音回复。

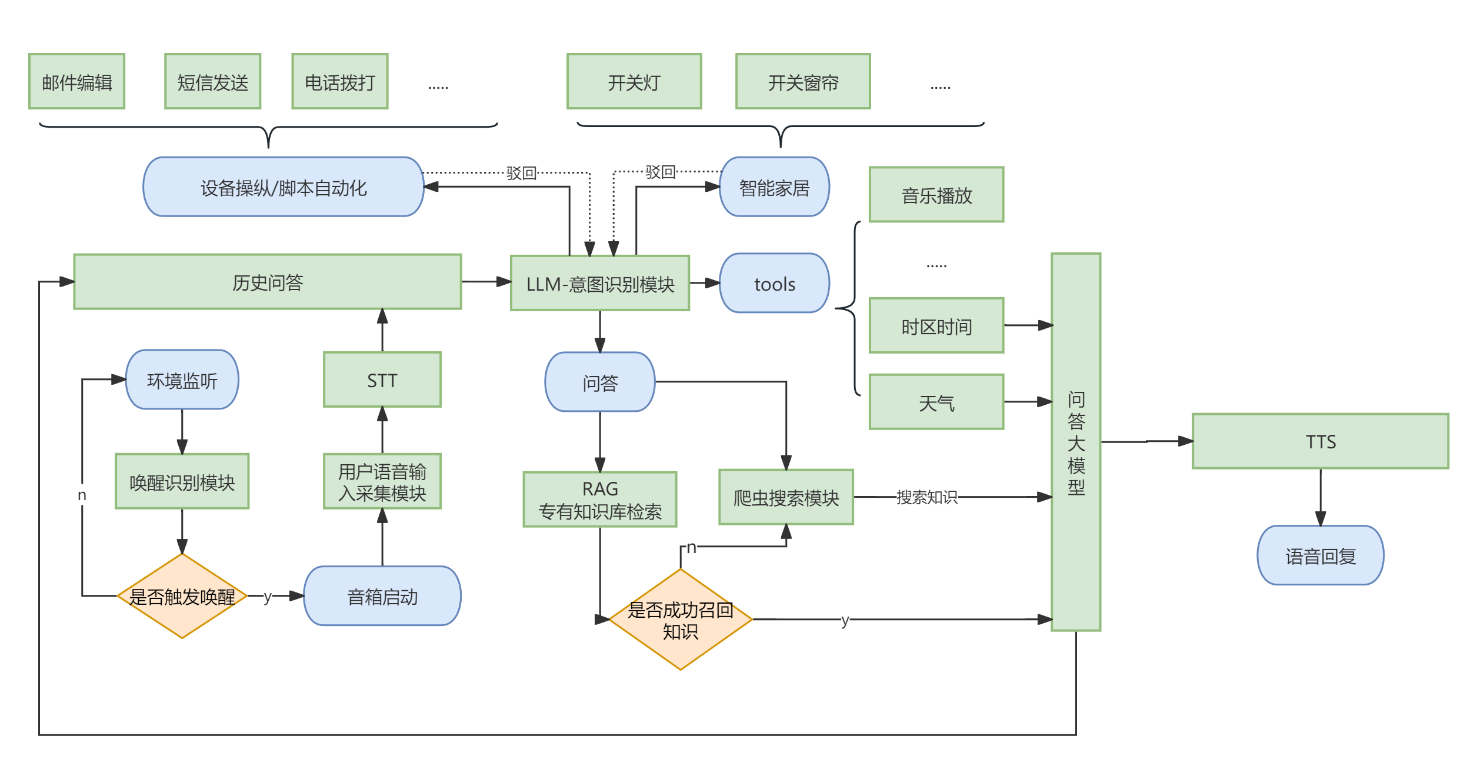
本发明的技术方案还包括一种基于Multi-Agent为核心的通用语音交互助手，其中包括9个主要功能模块，4个模型，2个存储库：

图2

1. 唤醒识别模块：用于监听环境，在触发唤醒词时启动 Multi-Agent 核心。
2. 意图识别模块/意图识别大模型：用于分析用户的文本输入，并将其分类为不同的意图类别。该模块利用一个经过微调的大语言模型（LLM）即意图识别大模型，结合预设的向量库，通过加权平均的方式输出最终的意图分类结果。
3. 知识补充模块：用于根据意图识别结果，选择并获取补充知识。该模块包括 RAG（检索增强生成）和多引擎爬虫搜索两种方式，以确保高效的知识扩展。
4. 工具调用模块：用于解析并执行意图识别模块中生成的工具调用语句，完成特定工具的调用并获取相应的补充信息。
5. 智能家居模块/智能家居大模型：用于判断意图识别模块分发的任务是否属于智能家居控制范畴，并执行相应的家居设备操作。
6. 设备操纵模块/设备操纵大模型：用于分析用户的指令并操作指定的设备，如手机或电脑，完成特定功能的调用。
7. 问答大模型：用于接收并处理所有的多轮文本输入和补充信息，生成最终的文本输出，并实时交付给 TTS 模块进行语音生成。
8. 语音转文本模块：用于将用户的语音输入转换为文本，以便后续处理。
9. 文本转语音模块：用于将系统生成的文本输出转换为语音回复用户。
10. RAG知识库：存储用于检索增强生成的知识数据，支持系统的知识扩展功能。
11. 历史对话数据库存：存储用户的历史对话记录，用于支持多轮对话和上下文理解。

# 五、技术效果

本发明所达到的有益效果是：通过创新的Multi-Agent架构等优化，实现了载体自由的智能语音助手框架。利用大模型与预设向量库的加权意图分类算法，显著提升了用户任务的精确分发。同时，将智能家居与设备操纵模块设立为独立智能体，减轻意图识别模块的负担，提高了系统的智能性与准确性。该设计确保系统链路简洁高效，同时提供高度个性化的服务，适应多种设备和应用场景，提升了用户体验。

# 六、本发明的优点

通过对比现有体系，本发明优势：

1. **载体自由**：本专利提出的技术方案是一个基于 Multi-Agent 的语音助手通用框架，能够适应多种载体平台。该框架的灵活性使其可以在不同设备上轻松部署。此外，去除首尾语音处理环节后，该框架依然可以作为一个强大的文本处理系统使用，为各类应用提供更广泛的支持。
2. **创新意图分类算法**：专利中引入了一种创新的意图分类算法，通过将大模型的分类结果与预设向量库的相似度进行加权平均，显著提高了意图识别的精度。该算法的应用确保了用户任务能够更准确地进行分发和处理，从而提升了系统的整体效率和用户体验。
3. **智能家居模块的融入**：专利将智能家居模块设立为一个独立的智能体（Agent），使系统能够自主判断和执行与智能家居相关的任务。智能家居模块结合意图识别模块，能独立分析用户的多轮对话输入，精准识别用户需求，智能调度家居设备，提供高度个性化的服务。
4. **设备操纵模块的集成**：专利中还集成了设备操纵模块，并将其设立为独立的智能体（Agent），增强了系统的设备控制能力。该模块通过对用户输入的独立分析，能够快速识别并执行涉及多种设备（如手机、电脑等）的操控任务。此功能不仅丰富了系统的交互方式，还大幅提高了用户操作的便捷性和效率。
5. **智能体架构的优化**：将智能家居和设备操纵模块分别设立为独立的智能体（Agent），减轻了意图识别模块的压力，提高了对用户意图的理解和处理的准确性与智能化。同时，这种架构可以校验意图识别模块的划分是否正确，使意图识别模块只需处理大领域的划分，而各小领域（如智能家居模块）内的任务执行则由相应模块负责分析和执行，增强了系统的智能性和准确性。
6. **保持链路简洁**：该架构在提升用户意图识别准确率的同时，没有显著增加系统链路的复杂度。对于每条链路的长度几乎相同，且较为简短，确保了对用户的及时响应。

# 七、本发明的关键点和保护点

1. 一种基于Multi-Agent为核心的通用语音交互助手架构，其特征在于：

步骤1：系统载体持续循环监听环境，并将监听到的内容传送至唤醒识别模块进行解析。一旦识别到用户设定的唤醒词，系统将发出语音交互提示，引导用户输入，并启动 Multi-Agent 核心。

步骤2：采集用户的语音输入，并根据预设的空白阈值对输入中的空白部分进行过滤。随后，调用 STT（Speech-to-Text，语音转文本）模块将用户的语音输入转换为文本。

步骤3：将前述文本输入存入历史对话数据库，并与历史对话数据一起提交给意图识别模块进行意图分类。根据分类结果，系统将决定是将任务分配给其他代理、调用工具获取扩展信息，还是进入问答流程。

步骤4：根据意图识别模块的反馈结果，选择适当的知识补充方式，将重构后的多轮文本输入与补充后的知识一并转交给步骤8的问答大模型进行生成输出。

步骤5：根据意图识别模块的反馈结果，调用指定工具获取补充信息，并将这些信息与多轮文本输入打包成一个集合，交给步骤8的问答大模型进行生成输出。

步骤6：智能家居模块会判断意图识别模块的任务派发是否合理，将不合理的派发驳回，对合理的任务派发进行进一步分析，并调用指定家居设备。

步骤7：设备操纵模块会判断意图识别模块的任务派发是否合理，将不合理的派发驳回，对合理的任务派发进行进一步分析，并操纵设备完成指定操作。

步骤8：将上述各步增补好知识的输入或者被驳回的多轮文本输入输入到问答大模型中，流式生成文本输出，并实时交付给 TTS（Text to Speech，文本转语音）模块生成语音回复。

2.所述步骤3具体包括两个步骤和四种可能触发结果：

1. 文本输入重构为多轮文本输入。
2. 调用意图识别模块对用户输入进行分发，分发给其它Agent或模块。

具体的计算方式如下：

1）调用微调后的意图分类大模型对重构后的多轮文本输入进行分析，计算每个场景的可能性。

|  |  |
| --- | --- |
|  | （2） |

其中即表示不同场景的可能性向量，场景包括智能家居（Smart Home）、设备操纵（Device Control）、工具调用（Tools）以及问答（QA）。每个维度的值表示该场景的可能性排名，例如，如果本轮输入属于智能家居的可能性排名第二则。

2）针对每个类别的可能用法构造数据，转化为向量并存储在向量库中。再将用户文本输入也转化为向量，并与每个类别的预设向量库计算相似度，得出一个可能性排名即，这个过程即。

|  |  |
| --- | --- |
|  | （3） |

3）对大模型分类结果和预设向量库相似度计算结果进行加权平均得到最终分类结果**。**

|  |  |
| --- | --- |
|  | （4） |

其中，权重的计算权重和的计算权重可以根据经验预设，也可以根据分类表现进行动态调整。

所述步骤3具体可能触发结果有四种，：

Ⅰ.若最终分类结果为问答QA，即,则跳转步骤4。

表示场景问答QA的加权排名值小于其它所有场景，即场景问答（QA）的可能性最高。下列公式相近，不再赘述。

Ⅱ.若最终分类结果为工具调用（Tools），即,则跳转步骤5。

Ⅲ.若最终分类结果为智能家居（Smart Home），即,则跳转步骤6。

Ⅳ.若最终分类结果为设备操纵（Device Control），即,则跳转步骤7。

值得注意的是，根据内部机制，参数和参数不会相同，因此最小值不会出现相等的情况。

3. 所述步骤4具体包括2种知识补充方式：

意图识别模块中的大模型在计算类别时，会提前判断选取何种知识补充方式，并通过 标记位指示是否使用 RAG 方式。如果最终分类场景为问答（QA）且，则启用 RAG 方式进行知识补充。否则，将通过多引擎爬虫搜索补充知识。此外，当 RAG 方式无法召回时，也会切换到多引擎爬虫搜索方式。

4.步骤6智能家居模块以Agent为核心，具有分析能力、驳回机制和智能家居执行能力：

（1）智能家居模块本身即为一个以大模型为核心的 Agent 智能体，因此不依赖意图识别模块的分析，可以自行对多轮文本输入 进行分析，判断其是否属于智能家居范畴。智能家居模块会将不合理的派发驳回，对合理的任务派发进行进一步分析，并调用指定家居设备。

如果属于智能家居范畴，则解析用户需要调用的家居设备及执行的功能，并将用户需求转化为智能家居调用语句并执行，其流程与 Tools 模块类似，封装为。若不属于智能家居范畴，则进行驳回。

（2）驳回有两种方式，对于时间敏感型场景，可以直接驳回给问答大模型，以便快速回复用户；对于时间相对不敏感的场景，则驳回给意图识别模块重新定向。系统最多只能重定向一次，若重定向失败则驳回给问答大模型。

5. 步骤7设备操纵模块以Agent为核心，具有分析能力、驳回机制和设备操纵脚本调用能力：

（1）操纵模块也是以大模型为核心的 Agent 智能体，因此不依赖意图识别模块的分析，可以自行对多轮文本输入进行分析，判断其是否属于设备操纵范畴。

如果属于设备操纵范畴，则解析用户需要调用的设备（如手机、电脑）及执行的功能，并将用户需求转化为设备调用语句并执行，流程与 Tools 模块类似，封装为。驳回机制与智能家居模块相同。

（2）驳回机制同智能家居模块。

6.本发明的技术方案还包括一种基于Multi-Agent为核心的通用语音交互助手，其中包括9个主要功能模块，4个模型，2个存储库：

1. 唤醒识别模块：用于监听环境，在触发唤醒词时启动 Multi-Agent 核心。
2. 意图识别模块/意图识别大模型：用于分析用户的文本输入，并将其分类为不同的意图类别。该模块利用一个经过微调的大语言模型（LLM）即意图识别大模型，结合预设的向量库，通过加权平均的方式输出最终的意图分类结果。
3. 知识补充模块：用于根据意图识别结果，选择并获取补充知识。该模块包括 RAG（检索增强生成）和多引擎爬虫搜索两种方式，以确保高效的知识扩展。
4. 工具调用模块：用于解析并执行意图识别模块中生成的工具调用语句，完成特定工具的调用并获取相应的补充信息。
5. 智能家居模块/智能家居大模型：用于判断意图识别模块分发的任务是否属于智能家居控制范畴，并执行相应的家居设备操作。
6. 设备操纵模块/设备操纵大模型：用于分析用户的指令并操作指定的设备，如手机或电脑，完成特定功能的调用。
7. 问答大模型：用于接收并处理所有的多轮文本输入和补充信息，生成最终的文本输出，并实时交付给 TTS 模块进行语音生成。
8. 语音转文本模块：用于将用户的语音输入转换为文本，以便后续处理。
9. 文本转语音模块：用于将系统生成的文本输出转换为语音回复用户。
10. RAG知识库：存储用于检索增强生成的知识数据，支持系统的知识扩展功能。
11. 历史对话数据库存：存储用户的历史对话记录，用于支持多轮对话和上下文理解。

# 摘要

本发明涉及一种基于大语言模型的多智能体语音交互助手系统，利用多个智能代理（Agent）来专注处理特定类型的任务或信息，并通过高效的协作机制实现更系统、更快速、更准确的语音理解和响应。具体步骤包括：1）监听与唤醒识别。系统循环监听环境中的语音信息。一旦检测到用户设定的唤醒词，系统会触发 Multi-Agent 核心大脑，并启动语音信息采集过程。2）语音转文本处理。对用户的语音输入进行转文本处理。用户的语音被转换为文本后，系统会进一步对文本中的空白部分进行过滤处理。3）意图识别与任务分发。系统将转化后的文本输入提交给意图识别模块进行分类。意图识别模块的任务是将用户输入分发到相应的智能代理或功能模块。4）智能家居模块的任务判断与调用。智能家居模块将对意图识别模块分发的任务进行判断。如果任务合理，则调用相应的家居服务；如果任务不合理，则进行驳回。5）设备操纵模块的任务判断与调用。设备操纵模块将对意图识别模块分发的任务进行判断。如果任务合理，则调用相应的设备操作服务；如果任务不合理，则进行驳回。6）处理意图识别反馈结果。根据意图识别模块的反馈结果，重构多轮文本输入，并将其与相关知识或工具调用结果整合后，交给大模型生成最终输出。7）生成最终输出并生成语音回复。大模型生成流式文本输出，并将其交给 TTS（文本转语音）模块生成语音回复，实时反馈给用户。