



(12) 发明专利申请

(10) 申请公布号 CN 120089140 A

(43) 申请公布日 2025. 06. 03

(21) 申请号 202510261979.6

G06F 40/253 (2020.01)

(22) 申请日 2025.03.05

G06F 40/30 (2020.01)

(71) 申请人 广州小鹏汽车科技有限公司

地址 510000 广东省广州市天河区岑村松
岗大街8号(72) 发明人 王小平 赵群 李晓辰 支淑婷
孟菲 王文杰(74) 专利代理机构 北京清亦华知识产权代理事
务所(普通合伙) 11201

专利代理师 魏宇晴

(51) Int.Cl.

G10L 15/26 (2006.01)

G10L 15/02 (2006.01)

G10L 15/18 (2013.01)

G06F 40/211 (2020.01)

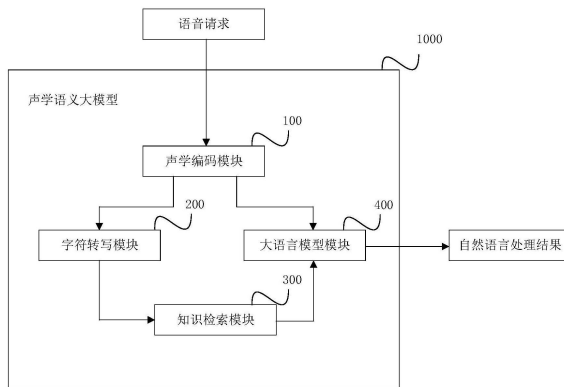
权利要求书2页 说明书13页 附图4页

(54) 发明名称

声学语义大模型、服务器、语音交互方法和
存储介质

(57) 摘要

本申请公开了一种声学语义大模型、服务器、语音交互方法和计算机可读存储介质。声学语义大模型包括声学编码模块、字符转写模块、知识检索模块和大语言模型模块。声学编码模块被配置为根据输入的语音请求,生成语音请求的声学特征向量。字符转写模块被配置为将语音请求转写为相对应的字符序列,字符序列包括语音请求中各文字相对应的字符。知识检索模块被配置为根据字符序列,自外部知识库中,获取补充信息。大语言模型模块被配置为根据声学特征向量和补充信息,确定自然语言处理结果。如此,通过端到端的声学语义大模型,减少了多个模块的串行处理,降低了处理语音请求的时延,提升了模型响应速度,从而增强用户体验。



1. 一种声学语义大模型,其特征在于,所述声学语义大模型包括声学编码模块、字符转写模块、知识检索模块和大语言模型模块;

所述声学编码模块被配置为根据输入的语音请求,生成所述语音请求的声学特征向量;

所述字符转写模块被配置为将所述语音请求转写为相对应的字符序列,所述字符序列包括所述语音请求中各文字相对应的字符;

所述知识检索模块被配置为根据所述字符序列,自外部知识库中,获取补充信息;

所述大语言模型模块被配置为根据所述声学特征向量和所述补充信息,确定自然语言处理结果。

2. 根据权利要求1所述的声学语义大模型,其特征在于,所述声学编码模块被配置为:

对所述语音请求进行特征提取处理,生成语义声学特征;

对所述语义声学特征进行编码处理,生成所述声学特征向量。

3. 根据权利要求1所述的声学语义大模型,其特征在于,所述字符转写模块被配置为:

基于预设算法,对所述声学特征向量进行映射处理,以将所述语音请求转写为所述字符序列。

4. 根据权利要求1所述的声学语义大模型,其特征在于,所述外部知识库基于键信息和值信息构建,其中,所述键信息基于汉语拼音和英文字母构建,所述值信息基于汉字构建;

所述外部知识库包括基础知识库,并以预定周期对所述基础知识库进行迭代更新。

5. 根据权利要求4所述的声学语义大模型,其特征在于,所述字符包括汉语拼音和/或英文字母,所述知识检索模块被配置为:

根据所述字符序列,自所述外部知识库中,确定与所述字符序列相对应的候选键信息;

根据所述候选键信息,将与所述候选键信息相对应的候选值信息,确定为所述候选信息;

根据所述候选信息的引用次数排序结果,自所述候选信息中,确定所述补充信息。

6. 根据权利要求1所述的声学语义大模型,其特征在于,所述大语言模型模块被配置为:

对所述声学特征向量和所述补充信息进行融合处理,生成融合信息;

根据所述融合信息,确定所述自然语言处理结果。

7. 根据权利要求6所述的声学语义大模型,其特征在于,所述大语言模型模块被配置为:

对所述融合信息进行槽位识别,得到槽位识别结果;

对所述融合信息进行应用程序接口预测,得到预测应用接口;

根据所述槽位识别结果和所述预测应用接口、选择所述预测应用接口执行应用程序接口参数填充,得到所述自然语言处理结果。

8. 一种服务器,其特征在于,所述服务器包括处理器和存储器,所述存储器上存储有计算机程序,当所述计算机程序被所述处理器执行时,实现权利要求1-7任一项所述的声学语义大模型。

9. 一种语音交互方法,其特征在于,所述语音交互方法基于如权利要求1-7任意一项所述的声学语义大模型,所述方法包括:

获取当前语音请求；

基于所述声学语义大模型,根据所述当前语音请求,确定与所述当前语音请求相对应的车辆控制指令,进行所述语音交互。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行的情况下,实现如权利要求9所述的方法的步骤。

声学语义大模型、服务器、语音交互方法和存储介质

技术领域

[0001] 本申请涉及语音交互领域,更具体而言,涉及一种声学语义大模型、服务器、语音交互方法和计算机可读存储介质。

背景技术

[0002] 相关技术中,为方便用户驾驶车辆,车载智能助手通常通过多个模块对语音请求的串行处理,包括通过声学拒识模块对语音请求进行声学拒识处理、通过语音识别模块对语音请求进行语音识别处理、基于知识检索模块对语音请求进行检索处理和利用自然语言处理模块对语音请求进行自然语言处理。然而如此,需要等待每个串行模块对语音请求的处理完成后,才能生成能够满足用户需求的自然语言理解结果,整体耗时较长,用户体验较差。

发明内容

[0003] 本申请提供了一种声学语义大模型、服务器、语音交互方法和计算机可读存储介质。

[0004] 本申请实施方式提供一种声学语义大模型,所述声学语义大模型包括声学编码模块、字符转写模块、知识检索模块和大语言模型模块;

[0005] 所述声学编码模块被配置为根据输入的语音请求,生成所述语音请求的声学特征向量;

[0006] 所述字符转写模块被配置为将所述语音请求转写为相对应的字符序列,所述字符序列包括所述语音请求中各文字相对应的字符;

[0007] 所述知识检索模块被配置为根据所述字符序列,自外部知识库中,获取补充信息;

[0008] 所述大语言模型模块被配置为根据所述声学特征向量和所述补充信息,确定自然语言处理结果。

[0009] 如此,声学语义大模型包括声学编码模块、字符转写模块、知识检索模块和大语言模型模块。其中,声学编码模块能够根据输入的语音请求,生成语音请求的声学特征向量。字符转写模块能够将语音请求转写为相对应的字符序列,字符序列包括语音请求中各文字相对应的字符。知识检索模块能够根据字符序列,自外部知识库中,获取补充信息。大语言模型模块能够根据声学特征向量和补充信息,确定自然语言处理结果。这样,通过端到端的声学语义大模型,减少了多个模块的串行处理,降低了处理语音请求的时延,提升了模型响应速度,从而增强用户体验。并且,通过汉语拼音或英文字母的匹配进行知识检索,避免了汉字解码的困难,提升了语音请求的识别准确性,从而增强声学语义大模型的语义理解能力和鲁棒性。

[0010] 某些实施方式中,所述声学编码模块被配置为:

[0011] 对所述语音请求进行特征提取处理,生成语义声学特征;

[0012] 对所述语义声学特征进行编码处理,生成所述声学特征向量。

[0013] 如此,声学编码模块能够对语音请求进行特征提取处理,生成语义声学特征。并能够对语义声学特征进行编码处理,生成声学特征向量。这样,通过提取语义声学特征,能够准确地捕捉语音信号的语义信息,并能够降低噪声等因素对语音识别的影响,提高声学语义大模型的语义理解能力和鲁棒性。

[0014] 在某些实施方式中,所述字符转写模块被配置为:

[0015] 基于预设算法,对所述声学特征向量进行映射处理,以将所述语音请求转写为所述字符序列。

[0016] 如此,字符转写模块能够基于预设算法,对声学特征向量进行映射处理,以将语音请求转写为字符序列。这样,通过预设算法对声学特征向量进行映射处理,能够准确地将语音信号转换为字符序列,为后续的知识检索提供基础。

[0017] 在某些实施方式中,所述外部知识库基于键信息和值信息构建,其中,所述键信息基于汉语拼音和英文字母构建,所述值信息基于汉字构建;

[0018] 所述外部知识库包括基础知识库,并以预定周期对所述基础知识库进行迭代更新。

[0019] 如此,外部知识库基于键信息和值信息构建,其中,键信息基于汉语拼音和英文字母构建,值信息基于汉字构建。此外,外部知识库包括基础知识库,并以预定周期对基础知识库进行迭代更新。这样,由于拼音和字母的索引方式比直接使用汉字更为简单和快速,通过使用汉语拼音和英文字母作为键信息,能够快速进行索引和检索。并且,通过预定周期对基础知识库进行迭代更新,可以确保外部知识库中内容的实时性。

[0020] 在某些实施方式中,所述字符包括汉语拼音和/或英文字母,所述知识检索模块被配置为:

[0021] 根据所述字符序列,自所述外部知识库中,确定与所述字符序列相对应的候选键信息;

[0022] 根据所述候选键信息,将与所述候选键信息相对应的候选值信息,确定为所述候选信息;

[0023] 根据所述候选信息的引用次数排序结果,自所述候选信息中,确定所述补充信息。

[0024] 如此,知识检索模块能够根据字符序列,自外部知识库中,确定与字符序列相对应的候选键信息。接着,知识检索模块能够根据候选键信息,将与候选键信息相对应的候选值信息,确定为候选信息。最后,知识检索模块能够根据候选信息的引用次数排序结果,自候选信息中,确定补充信息。这样,通过候选键信息和字符序列的匹配,可以确保检索到的补充信息与用户输入的字符序列相匹配,从而提高知识检索的准确性。并且,通过候选信息引用次数的排序,可以优先选择引用次数最高的信息,从而合理地处理多义字的情况。

[0025] 在某些实施方式中,所述大语言模型模块被配置为:

[0026] 对所述声学特征向量和所述补充信息进行融合处理,生成融合信息;

[0027] 根据所述融合信息,确定所述自然语言处理结果。

[0028] 如此,大语言模型模块能够对声学特征向量和补充信息进行融合处理,生成融合信息。接着,大语言模型模块能够根据融合信息,确定自然语言处理结果。这样,通过融合声学特征向量和补充信息,能够帮助大语言模型准确地理解用户意图,从而提高识别准确率,增强用户体验。

- [0029] 在某些实施方式中,所述大语言模型模块被配置为:
- [0030] 对所述融合信息进行槽位识别,得到槽位识别结果;
- [0031] 对所述融合信息进行应用程序接口预测,得到预测应用接口;
- [0032] 根据所述槽位识别结果和所述预测应用接口、选择所述预测应用接口执行应用程序接口参数填充,得到所述自然语言处理结果。
- [0033] 如此,大语言模型模块能够对融合信息进行槽位识别,得到槽位识别结果。接着,大语言模型模块能够对融合信息进行应用程序接口预测,得到预测应用接口。最后,大语言模型模块能够根据槽位识别结果和预测应用接口、选择预测应用接口执行应用程序接口参数填充,得到自然语言处理结果。这样,通过槽位识别,大语言模型能够准确地理解用户指令中各个实体的含义,从而提升用户体验。通过应用接口预测和参数填充,大语言模型能够根据用户指令选择合适的接口并填充相应的参数,实现灵活的指令处理,从而提升用户体验。
- [0034] 本申请实施方式提供了一种服务器,所述服务器部署有上述的声学语义大模型。
- [0035] 本申请实施方式提供一种语音交互方法,所述方法包括:
- [0036] 获取当前语音请求;
- [0037] 基于所述声学语义大模型,根据所述当前语音请求,确定与所述当前语音请求相对应的车辆控制指令,进行所述语音交互。
- [0038] 如此,获取当前语音请求。接着,基于声学语义大模型,根据当前语音请求,确定与当前语音请求相对应的车辆控制指令,进行语音交互。这样,基于上述的声学语义大模型,能够迅速且准确地对用户语音请求进行处理,增强用户体验。
- [0039] 本申请实施方式提供了一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行的情况下,实现如上述的语音交互方法的步骤。
- [0040] 本申请的实施方式的附加方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本申请的实施方式的实践了解到。

附图说明

- [0041] 本申请的上述和/或附加的方面和优点从结合下面附图对实施方式的描述中将变得明显和容易理解,其中:
- [0042] 图1是本申请实施方式的声学语义大模型的结构示意图;
- [0043] 图2是本申请实施方式的语音交互方法的流程示意图之一;
- [0044] 图3是本申请实施方式的语音交互方法的流程示意图之二;
- [0045] 图4是本申请实施方式的语音交互方法的流程示意图之三;
- [0046] 图5是本申请实施方式的语音交互方法的流程示意图之四;
- [0047] 图6是本申请实施方式的语音交互方法的流程示意图之五;
- [0048] 图7是本申请实施方式的语音交互方法的流程示意图之六;
- [0049] 图8是本申请实施方式的语音交互方法的流程示意图之七。

具体实施方式

- [0050] 下面详细描述本申请的实施方式,实施方式的示例在附图中示出,其中,相同或类

似的标号自始至终表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施方式是示例性的,仅用于解释本申请的实施方式,而不能理解为对本申请的实施方式的限制。

[0051] 相关技术中,为了方便用户在驾驶车辆时能够安全、有效地进行交互,往往使用车载智能助手辅助用户及时车辆。对于用户发出的语音请求,车载智能助手普遍采用多个模块对语音请求进行串行处理:

[0052] 首先,通过声学拒识模块对用户的语音请求进行声学拒识处理,声学拒识模型分析语音请求的音频质量,判断是否为噪声或受损音频。若语音请求的音频质量不良(如背景嘈杂或断断续续),则判定为无效语音请求,直接拒识,结束流程。若语音请求的音频质量良好,则判定为有效语音请求,放行语音请求至下一模块。

[0053] 接着,有效语音请求会被送入语音识别模块,语音识别模块能够将用户的语音请求转换为文字,即对语音信号进行解码,将其转换为可理解的文本信息。然而,若是用户语音请求表述不标准,语音识别模块可能无法将用户语音请求转换为正确的文本信息。例如,用户语音请求为“我想我妈了,帮我打电话给她”,语音识别模块有可能因为口音问题,将语音请求转换为“我想我马了,帮我打电话给她”。这样,后续步骤将基于错误文本“我想我马了,帮我打电话给她”进行,无法满足用户需求。

[0054] 然后,转换后的文本信息会进入知识检索模块,知识检索模块能够根据用户的请求,在车载智能助手的知识库中检索相关的信息。若检索到相关条目,将其作为补充知识。若未检索到(如无匹配内容或相似度低于阈值),则不添加额外知识。例如,如果用户询问附近的餐馆,知识检索模块会搜索用户当前的位置信息,并检索附近的餐馆信息。

[0055] 最后,检索到的信息会进入自然语言处理模块,自然语言处理模块能够理解文本信息和补充信息的含义,包括但不限于词法分析、句法分析、语义理解等,最终生成满足用户需求的自然语言理解结果。

[0056] 然而,上述串行处理流程需要依次等待每个模块完成处理,导致整体响应时间较长,用户体验较差,尤其是在需要快速响应的场景下,例如导航或紧急呼叫等。

[0057] 基于上述的问题,请参阅图1,本申请实施方式提供一种声学语义大模型1000,声学语义大模型1000包括声学编码模块100、字符转写模块200、知识检索模块300和大语言模型模块400。

[0058] 其中,声学编码模块100能够根据输入的语音请求,生成语音请求的声学特征向量;

[0059] 字符转写模块200能够将语音请求转写为相对应的字符序列,字符序列包括语音请求中各文字相对应的字符;

[0060] 知识检索模块300能够根据字符序列,自外部知识库中,获取补充信息;

[0061] 大语言模型模块400能够根据声学特征向量和补充信息,确定自然语言处理结果。

[0062] 具体地,声学语义大模型1000是一个端到端系统,通过将声学编码处理、拼音转写处理、知识检索处理和自然语言处理融合在一起,实现了端到端的处理,避免了传统方案中各个模块之间需要单独训练和调优的复杂性,从而提高了系统的响应速度。在某些实施方式中,声学语义大模型1000通常采用深度学习技术构建,例如卷积神经网络(CNN)、循环神经网络(RNN)或Transformer模型等,在此不做限定。声学语义大模型1000能够有效地捕捉

语音信号的复杂特征和语义信息,从而提高语音识别和语义理解的准确率。并且,声学语义大模型1000通过转写为字符序列,能够适应不同的说话人、口音和噪声环境,具有很强的鲁棒性。需要说明地,该声学语义大模型1000不仅能够部署在车载智能助中,在智能家居、智能客服等领域也有着广泛的应用前景。

[0063] 声学编码模块100是声学语义大模型1000中负责将语音信号转换为声学特征向量的模块,通过信号处理技术,例如傅里叶变换、滤波器组等,将语音信号转换成声学特征向量,为后续的任务提供数据信息支持。声学编码模块100相当于传统方案中声学拒识模块和语音识别模块的结合,能够实现声学拒识模块和语音识别模块的部分功能。这样,声学编码模块100将声学特征提取和语音识别融合在一起,实现了端到端的处理,从而提高了系统的响应速度。

[0064] 语音请求指的是用户通过语音方式向智能座舱系统发出的指令或提问,包括多种类型的信息。例如,例如,“播放歌手A的歌曲B”、“导航去北京”、“今天的天气怎么样”、“附近有什么餐厅”、“打开空调”和“讲个笑话”等。

[0065] 声学特征向量指的是从语音请求中提取的能够表征语音请求声学特性的多维向量,包括语音请求的重要信息,例如音高、音强和音色等,是语音识别、语音合成和说话人识别等语音处理任务的基础。

[0066] 字符转写模块200是声学语义大模型1000中将语音识别模型输出的声学特征向量转换为拼音序列的模块,为知识检索模块300提供输入,并提高大语言模型模块400的准确率。

[0067] 字符序列指的是由汉语拼音和/或英文字母等字符组成的序列,用于表示语音请求中的文字信息,可以用于检索与语音请求相关的知识。

[0068] 知识检索模块300是声学语义大模型1000中根据字符序列从外部知识库中检索相关知识的模块,能够从外部知识库中获取补充信息,并应用于后续模块。

[0069] 外部知识存储有各种知识信息,能够为知识检索模块300提供知识来源。

[0070] 补充信息指的是从外部知识库中检索到的与拼音序列相关的知识,为大语言模型模块400提供额外的知识,帮助其更好地理解输入文本,并做出更准确的响应。

[0071] 大语言模型模块400指的是声学语义大模型1000中根据声学特征向量和补充信息,确定自然语言处理结果的模块,能够理解和生成自然语言,并进行语义理解和意图识别,输出自然语言处理结果,如应用程序接口API和对应的填充参数。在某些实施方式中,大语言模型模块400中的大语言模型可以是如BERT等基于Transformer的模型,在此不做限定。

[0072] 首先,声学编码模块100接收输入的语音请求,并将语音请求转换为声学特征向量。接着,字符转写模块200将语音请求转写为相对应的字符序列,字符序列包括语音请求中各文字相对应的字符,即汉语拼音和/或英文字母。由于汉语拼音和英文字母的粒度较粗,解码准确性更高。然后,知识检索模块300根据拼音序列,从外部知识库中检索与用户输入相关的额外知识,例如歌曲信息、地名信息等。最后,大语言模型模块400结合声学特征向量和检索到的额外知识,理解用户意图并生成自然语言处理结果,例如API和Arguments。

[0073] 综上,本申请提供的声学语义大模型1000包括声学编码模块100、字符转写模块200、知识检索模块300和大语言模型模块400。其中,声学编码模块100能够根据输入的语音

请求,生成语音请求的声学特征向量。字符转写模块200能够将语音请求转写为相对应的字符序列,字符序列包括语音请求中各文字相对应的字符。知识检索模块300能够根据字符序列,自外部知识库中,获取补充信息。大语言模型模块400能够根据声学特征向量和补充信息,确定自然语言处理结果。这样,通过端到端的声学语义大模型1000,减少了多个模块的串行处理,降低了处理语音请求的时延,提升了模型响应速度,从而增强用户体验。并且,通过汉语拼音或英文字母的匹配进行知识检索,避免了汉字解码的困难,提升了语音请求的识别准确性,从而增强声学语义大模型1000的语义理解能力和鲁棒性。

[0074] 在某些实施方式中,声学编码模块100能够对语音请求进行特征提取处理,生成语义声学特征。并能够对语义声学特征进行编码处理,生成声学特征向量。

[0075] 具体地,特征提取处理指的是从原始语音请求中提取出与语音识别、语义理解等任务相关的有用信息的过程,包括噪声处理、归一化处理、帧处理,特征计算处理和特征归一化处理等。特征提取处理类似于人类听觉系统对声音的感知过程,即,从声音中提取出有用的信息,例如声音的音高、音量、音色等,并根据这些信息理解声音的含义。特征提取处理通过模拟人类听觉系统对声音的感知过程,将语音信号转换成计算机可以理解的特征,以便进行后续的语音识别、语义理解等任务。

[0076] 语义声学特征指的是从语音请求信号中提取出的与语义信息相关的声学特征,是连接语音信号和语义信息之间的桥梁。语义声学特征包括音素特征、音节特征和韵律特征等语音请求信号中与语义相关的信息。

[0077] 其中,音素特征指的是指与单个音素相关的声学特征,包括音素时长、音素能量和音素频率等,音素是语音学中用来描述语音的最小单位,它具有区别意义的作用。例如,汉语中的“妈”和“马”的发音区别在于“a”和“ā”的音素不同,前者是开口呼,后者是合口呼。音素时长指的是音素在语音中的持续时间。音素能量指的是音素的能量强度。音素频率指的是音素的频率成分。

[0078] 音节特征指的是指与单个音节相关的声学特征,包括音节时长、音节能量、音节时长分布和音节能量分布等,音节是语音结构的基本单位,由一个或多个音素组成,并包括一个核心音素。音节是语言中可以独立发音的最小单位,例如汉语中的“啊”和“衣”等都是一个音节。音节时长指的是音节在语音中的持续时间。音节能量指的是音节的能量强度。音节时长分布指的是音节内部各音素时长的分布情况。音节能量分布指的是音节内部各音素能量的分布情况。

[0079] 韵律特征指的是与语音韵律相关的声学特征,包括语调、语速和停顿等。语调指的是语音的升降变化,例如升调、降调、平调等。语速指的是语音的快慢程度。停顿指的是语音的停顿位置和停顿时间。

[0080] 以“播放歌手A的歌曲B”这个语音请求为例,语义声学特征包括以下内容:

[0081] 音素特征:例如“b”、“o”、“f”、“a”、“ng”等音素的时长、能量等。

[0082] 音节特征:例如“bo”、“fang”、“ge”、“shou”、“A”、“ge”、“qu”、“B”等音节的时长、能量和时长分布等。

[0083] 韵律特征:例如语调的升降、语速的快慢、停顿的位置等。

[0084] 声学特征向量指的是将语音请求信号转换成计算机可以理解的数学形式,包括语音信号中与语义相关的信息,可以用于后续的自然语言处理和知识检索等。声学特征向量

通常包括如梅尔频率倒谱系数特征值、音素特征值、音节特征值和韵律特征值等多个特征值。梅尔频率倒谱系数特征值 (Mel-frequency Cepstral Coefficients, MFCC) 是一种常用的语音特征参数,它通过将语音信号转换到梅尔频率域,并对对数谱进行离散余弦变换来提取特征。

[0085] 声学编码模块100首先对输入的语音请求进行特征提取,生成语义声学特征。接着,声学编码模块100将语义声学特征编码成声学特征向量。这些声学特征向量是高维空间中的点,能够代表语音信号的复杂特性。

[0086] 如此,声学编码模块100能够对语音请求进行特征提取处理,生成语义声学特征。并能够对语义声学特征进行编码处理,生成声学特征向量。这样,通过提取语义声学特征,能够准确地捕捉语音信号的语义信息,并能够降低噪声等因素对语音识别的影响,提高声学语义大模型1000的语义理解能力和鲁棒性。

[0087] 在某些实施方式中,字符转写模块200能够基于预设算法,对声学特征向量进行映射处理,以将语音请求转写为字符序列。

[0088] 具体地,字符转写模块200是一个预训练完成的模型,能够将声学特征向量映射为字符序列。在某些实施方式中,字符转写模块200能够将Recurrent Neural Network、Convolutional Neural Network和transformer深度神经网络模型其中之一作为基座模型。随后,根据预设训练数据对基座模型进行训练,并选择相应预设算法,构建得到字符转写模块200。需要说明地,根据预设训练数据对基座模型进行训练时,主要学习将声学特征向量映射为汉语拼音和如何将声学特征向量映射为26个英文字母。相比于使用汉字进行训练,通过使用拼音和英文字母进行训练,字符转写模块200的建模的粒度更粗、建模难度更低,并且训练完成的字符转写模块200解码准确性更高。例如,用户语音请求“我想我妈(由于地域性的口音,说成了第三声)了,帮我打电话给她”,若是直接转为汉字可能就是“我想我马了,帮我打电话给她”,影响后续处理。而转为汉语拼音“wo xiang wo ma le, bang wo da dian hua gei ta”,在后续大语言模型模块400进行处理时,能够发现并纠正。

[0089] 此外,使用拼音和英文字母进行建模,可以在外部知识库中根据拼音和英文字母进行知识匹配,例如可以匹配“ge shou A”和“ge qu B”等拼音序列,从而提高知识检索的准确性。并且,使用拼音和英文字母进行建模,可以更好地处理新词、热词和地名,即使模型训练时没有见过这些词,也能够通过拼音进行匹配,例如可以解码出“ge qu B”等新词,从而提高字符转写模块200的鲁棒性。

[0090] 预设算法指的是字符转写模块200中的解码算法,用于指导字符序列转写的过程。在某些实施方式中,预设算法能够是贪心算法、Beam Search算法、Prefix Beam Search算法和Modified Beam Search算法中的任意一种,在此不做限定。在本申请实施方式所举示例中,考虑到声学语义大模型1000的时延要求,使用贪心算法作为预设算法。贪心算法一种启发式搜索算法,通过在每一步选择当前看起来最优的选项,来尝试找到问题的最优解。

[0091] 拼音转写模块接收声学编码模块100输出的声学特征向量。接着,拼音转写模块使用预设算法对声学特征向量进行映射处理,将声学特征向量转换为字符序列,这些字符序列对应于输入声学语义大模型1000的语音请求。

[0092] 如此,字符转写模块200能够基于预设算法,对声学特征向量进行映射处理,以将语音请求转写为字符序列。这样,通过预设算法对声学特征向量进行映射处理,能够准确地

将语音信号转换为字符序列,为后续的知识检索提供基础。

[0093] 在某些实施方式中,外部知识库基于键信息和值信息构建,其中,键信息基于汉语拼音和英文字母构建,值信息基于汉字构建。并且,外部知识库包括基础知识库,并以预定周期对基础知识库进行迭代更新。

[0094] 具体地,外部知识库基于键信息和值信息构建。其中,键信息(Key)由汉语拼音和英文字母构建,这意味着外部知识库中的每个条目都通过拼音和字母进行索引,而不是直接使用汉字,相比于直接使用汉字进行匹配,拼音匹配可以更好地应对新词、热词和地名等变化,例如用户可以说“qi che shou ce”,外部知识库能够找到“汽车使用手册”的相关信息。值信息(Value)由汉字构建,即具体的知识内容,使用汉字进行描述,每个键信息都对应一个或多个值信息。在某些实施方式中,外部知识库的构建流程如下:第一,进行数据收集,从互联网、书籍、百科全书等来源收集知识数据,例如歌词、电影台词、新闻等。第二,进行数据清洗,对收集到的数据进行清洗,去除噪声、重复信息等,并进行分词和词性标注。第三,将清洗后的数据构建键值对形式,其中键信息基于汉语拼音和英文字母构建,值信息基于汉字构建。例如,将“汽车”的键信息设置为“qi che”,值信息设置为“汽车是一种代步工具”。第四,知识库构建,将构建好的键值对存储在外部知识库中,并按照拼音字母顺序进行组织。

[0095] 外部知识库以基础知识库作为外部知识库的主体部分,基础知识库中包括大量的基础知识和信息。并且,基础知识库会定期进行更新,以保证知识的时效性和准确性,更好地满足用户的需求。更新周期是预定的,可以是每天、每周、每月等。

[0096] 如此,外部知识库基于键信息和值信息构建,其中,键信息基于汉语拼音和英文字母构建,值信息基于汉字构建。此外,外部知识库包括基础知识库,并以预定周期对基础知识库进行迭代更新。这样,由于拼音和字母的索引方式比直接使用汉字更为简单和快速,通过使用汉语拼音和英文字母作为键信息,能够快速进行索引和检索。并且,通过预定周期对基础知识库进行迭代更新,可以确保外部知识库中内容的实时性。

[0097] 在某些实施方式中,知识检索模块300能够根据字符序列,自外部知识库中,确定与字符序列相对应的候选键信息。

[0098] 接着,知识检索模块300能够根据候选键信息,将与候选键信息相对应的候选值信息,确定为候选信息。

[0099] 最后,知识检索模块300能够根据候选信息的引用次数排序结果,自候选信息中,确定补充信息。

[0100] 具体地,候选键信息是与用户输入的字符序列相匹配的信息,可以作为检索知识库的线索,即“索引”或“关键词”,能够帮助我们缩小检索范围,快速找到与用户输入相关的信息。例如,输入字符序列为“qi che”,候选键信息可能为“qi che an quan”、“qi che jia shi qing kuang”和“qi che bao yang qing kuang”等。

[0101] 候选值信息是指与候选键信息相对应的,从外部知识库中提取出来的信息。候选值信息是与输入的字符序列相关的,可以帮助用户获取更全面的信息,即“详细信息”,提供了与用户输入相关的更具体的信息。

[0102] 候选信息指的知识检索过程中,从外部知识库中检索到的与输入的字符序列相关的所有知识片段。

[0103] 候选信息的引用次数指的是每个候选信息在知识库中被引用的次数,可以理解为该候选信息的重要程度或热度。引用次数排序结果指的是根据候选信息的引用次数,将这些候选信息按照重要程度或热度进行排序。例如,输入的字符序列为“ge shou A”进行知识检索,知识库中检索到了以下三个与“ge shou A”相关的候选信息:候选信息A:ge shou A,著名歌手,代表作有《ge qu B》、《ge qu C》等。候选信息B:ge shou A,知名演员,代表作有《ying shi D》、《ying shi F》等。候选信息C:ge shou A,著名篮球运动员,效力于知名球队。假设候选信息A的引用次数最高,候选信息B的引用次数次之,候选信息C的引用次数最低。那么,根据候选信息的引用次数排序结果,排名由高到低,这三个候选信息将被排序为:候选信息A、候选信息B和候选信息C。这样,通过候选信息的排序,可以优先选择引用次数最高的信息,从而更好地处理多义字的情况。

[0104] 需要说明地,知识检索模块300采用检索增强生成技术(Retrieval Augmented Generation,RAG),RAG知识检索技术是一种结合信息检索和自然语言生成的技术,旨在利用外部知识库来增强模型的性能。

[0105] 如此,知识检索模块300能够根据字符序列,自外部知识库中,确定与字符序列相对应的候选键信息。接着,知识检索模块300能够根据候选键信息,将与候选键信息相对应的候选值信息,确定为候选信息。最后,知识检索模块300能够根据候选信息的引用次数排序结果,自候选信息中,确定补充信息。这样,通过候选键信息和字符序列的匹配,可以确保检索到的补充信息与用户输入的字符序列相匹配,从而提高知识检索的准确性。并且,通过候选信息引用次数的排序,可以优先选择引用次数最高的信息,从而合理地处理多义字的情况。

[0106] 在某些实施方式中,大语言模型模块400能够对声学特征向量和补充信息进行融合处理,生成融合信息。

[0107] 接着,大语言模型模块400能够根据融合信息,确定自然语言处理结果。

[0108] 具体地,融合处理指的是将声学特征向量和补充信息结合起来,以便大语言模型模块400更好地理解 and 处理用户语音请求。在某些实施方式中,融合处理的方式包括特征拼接和特征加权拼接等,特征拼接指的是将声学特征向量和补充信息直接拼接成一个更长的特征向量。特征加权拼接指的是对声学特征向量和补充信息进行加权,然后进行拼接。

[0109] 将声学特征向量和补充信息输入到大语言模型模块400,该模块会对这些信息进行处理,结合声学特征向量和补充信息,生成语义理解更加全面的融合信息。声学特征向量包含了丰富的声学信息,补充信息则提供了额外的语义信息,两者的融合能够帮助大语言模型更准确地理解用户意图,从而提高识别准确率。

[0110] 接着,大语言模型模块400根据融合信息,确定自然语言处理结果,例如识别用户指令、回答用户问题等。

[0111] 如此,大语言模型模块400能够对声学特征向量和补充信息进行融合处理,生成融合信息。接着,大语言模型模块400能够根据融合信息,确定自然语言处理结果。这样,通过融合声学特征向量和补充信息,能够帮助大语言模型准确地理解用户意图,从而提高识别准确率,增强用户体验。

[0112] 在某些实施方式中,大语言模型模块400能够对融合信息进行槽位识别,得到槽位识别结果。

[0113] 接着,大语言模型模块400能够对融合信息进行应用程序接口预测,得到预测应用接口。

[0114] 最后,大语言模型模块400能够根据槽位识别结果和预测应用接口、选择预测应用接口执行应用程序接口参数填充,得到自然语言处理结果。

[0115] 具体地,槽位识别指的是从用户的输入中提取特定的信息片段,这些信息片段通常被称为“槽位”(slots)。槽位通常是完成某个任务或请求所必需的关键信息,如时间、地点、对象等。以用户语音请求为“明天温度多少”为例,进行槽位识别可以得到的槽位信息包括[“明天”——日期(Date)],即槽位信息包括槽位取值和槽位类型,其中“明天”为槽位取值,日期(Date)为槽位类型。以用户语音请求“导航到地址Q”为例,进行槽位识别可以得到的槽位信息为[“地址Q”——地名(Place)],其中“中关村”为槽位取值,地名(Place)为槽位类型。槽位识别是自然语言处理中的一个重要任务,其目标是识别文本中的特定实体和属性,并将其与预定义的槽位对应起来。槽位识别可以帮助大语言模型模块400理解用户的意图,并生成更准确、更自然的回复。

[0116] 槽位识别结果指的是对语音请求进行槽位识别,得到的命名实体。如上述的槽位信息[“明天”——日期(Date)]和槽位信息[“地址Q”——地名(Place)]等。

[0117] 应用程序接口预测指的是根据输入文本的语义,预测出与输入文本对应的操作类型,并生成相应的应用程序接口调用指令。应用程序接口预测可以帮助大语言模型模块400理解用户的意图,并执行相应的操作。

[0118] 应用程序接口参数填充是自然语言处理中的一个任务,其目标是根据输入文本的语义和应用程序接口调用指令,为应用程序接口调用指令中的参数指定具体的值。

[0119] 大语言模型模块400首先对融合信息进行槽位识别,将用户的指令分解为不同的槽位,例如歌曲名、歌手名和播放模式等。接着,大语言模型模块400根据融合信息进行应用程序接口预测,预测用户需要执行的应用程序接口,例如“播放歌曲”和“查询天气”等。最后,大语言模型模块400根据槽位识别结果和预测应用接口,选择相应的预测应用接口执行应用程序接口参数填充,例如将歌曲名和歌手名作为参数填充到“播放歌曲”接口中。

[0120] 最终,大语言模型模块400输出自然语言处理结果,例如识别用户指令为“播放歌曲”,并调用音乐播放器播放用户指定的歌曲。

[0121] 如此,大语言模型模块400能够对融合信息进行槽位识别,得到槽位识别结果。接着,大语言模型模块400能够对融合信息进行应用程序接口预测,得到预测应用接口。最后,大语言模型模块400能够根据槽位识别结果和预测应用接口、选择预测应用接口执行应用程序接口参数填充,得到自然语言处理结果。这样,通过槽位识别,大语言模型能够准确地理解用户指令中各个实体的含义,从而提升用户体验。通过应用接口预测和参数填充,大语言模型能够根据用户指令选择合适的接口并填充相应的参数,实现灵活的指令处理,从而提升用户体验。

[0122] 本申请实施方式提供了一种服务器,服务器部署有上述的声学语义大模型1000。

[0123] 具体地,服务器是运行应用程序并提供服务的硬件设备,能够为声学语义大模型1000提供运行环境,包括计算资源、存储空间和网络连接等。并能够将训练好的声学语义大模型1000部署到线上环境,使其能够对外提供服务。需要说明地,车辆的算力足够声学语义大模型1000使用时,服务器能够部署在车辆本地。反之,则部署在云端。

[0124] 服务器接收来自用户的语音输入,并将处理结果返回给用户。

[0125] 如此,服务器将声学语义大模型1000集成到一个完整的系统中,实现从语音输入到自然语言处理结果输出的端到端流程,快速响应用户的语音指令,并提供准确的结果,从而提升用户体验。

[0126] 请参阅图2,本申请实施方式提供一种语音交互方法,方法包括:

[0127] 01:获取当前语音请求;

[0128] 02:基于声学语义大模型,根据当前语音请求,确定与当前语音请求相对应的车辆控制指令,进行语音交互。

[0129] 具体地,系统通过麦克风等硬件设备接收用户发出的当前语音请求。接着,声学语义大模型1000对当前语音请求进行理解和分析,识别出用户的意图和目标,并将其转换为车辆可以理解的车辆控制指令。

[0130] 请参阅图3,在某些实施方式中,步骤02(基于声学语义大模型,根据当前语音请求,确定与当前语音请求相对应的车辆控制指令),包括:

[0131] 021:根据输入的语音请求,生成语音请求的声学特征向量;

[0132] 022:将语音请求转写为相对应的字符序列;

[0133] 023:根据字符序列,自外部知识库中,获取补充信息;

[0134] 024:根据声学特征向量和补充信息,确定自然语言处理结果。

[0135] 如此,通过端到端的声学语义大模型,减少了多个模块的串行处理,降低了处理语音请求的时延,提升了模型响应速度,从而增强用户体验。并且,通过汉语拼音或英文字母的匹配进行知识检索,避免了汉字解码的困难,提升了语音请求的识别准确性,从而增强声学语义大模型的语义理解能力和鲁棒性。

[0136] 请参阅图4,在某些实施方式中,步骤021(根据输入的语音请求,生成语音请求的声学特征向量),包括:

[0137] 0211:对语音请求进行特征提取处理,生成语义声学特征;

[0138] 0212:对语义声学特征进行编码处理,生成声学特征向量。

[0139] 如此,通过提取语义声学特征,能够准确地捕捉语音信号的语义信息,并能够降低噪声等因素对语音识别的影响,提高声学语义大模型的语义理解能力和鲁棒性。

[0140] 请参阅图5,在某些实施方式中,步骤022(将语音请求转写为相对应的字符序列),包括:

[0141] 0221:基于预设算法,对声学特征向量进行映射处理,以将语音请求转写为字符序列。

[0142] 如此,通过预设算法对声学特征向量进行映射处理,能够准确地将语音信号转换为字符序列,为后续的知识检索提供基础。

[0143] 在某些实施方式中,外部知识库基于键信息和值信息构建,其中,键信息基于汉语拼音和英文字母构建,值信息基于汉字构建。并且,外部知识库包括基础知识库,并以预定周期对基础知识库进行迭代更新。

[0144] 如此,由于拼音和字母的索引方式比直接使用汉字更为简单和快速,通过使用汉语拼音和英文字母作为键信息,能够快速进行索引和检索。并且,通过预定周期对基础知识库进行迭代更新,可以确保外部知识库中内容的实时性。

[0145] 请参阅图6,在某些实施方式中,字符包括汉语拼音和/或英文字母,步骤023(根据字符序列,自外部知识库中,获取补充信息),包括:

[0146] 0231:根据字符序列,自外部知识库中,确定与字符序列相对应的候选键信息;

[0147] 0232:根据候选键信息,将与候选键信息相对应的候选值信息,确定为候选信息;

[0148] 0233:根据候选信息的引用次数排序结果,自候选信息中,确定补充信息。

[0149] 如此,通过候选键信息和字符序列的匹配,可以确保检索到的补充信息与用户输入的字符序列相匹配,从而提高知识检索的准确性。并且,通过候选信息引用次数的排序,可以优先选择引用次数最高的信息,从而合理地处理多义字的情况。

[0150] 请参阅图7,在某些实施方式中,步骤024(根据声学特征向量和补充信息,确定自然语言处理结果),包括:

[0151] 0241:对声学特征向量和补充信息进行融合处理,生成融合信息;

[0152] 0242:根据融合信息,确定自然语言处理结果。

[0153] 如此,通过融合声学特征向量和补充信息,能够帮助大语言模型准确地理解用户意图,从而提高识别准确率,增强用户体验。

[0154] 请参阅图8,在某些实施方式中,步骤0242(根据融合信息,确定自然语言处理结果),包括:

[0155] 02421:对融合信息进行槽位识别,得到槽位识别结果;

[0156] 02422:对融合信息进行应用程序接口预测,得到预测应用接口;

[0157] 02423:根据槽位识别结果和预测应用接口、选择预测应用接口执行应用程序接口参数填充,得到自然语言处理结果。

[0158] 如此,通过槽位识别,大语言模型能够准确地理解用户指令中各个实体的含义,从而提升用户体验。通过应用接口预测和参数填充,大语言模型能够根据用户指令选择合适的接口并填充相应的参数,实现灵活的指令处理,从而提升用户体验。

[0159] 需要说明地,本申请实施提供的语音交互方法基于前述声学语义大模型1000实现。具体而言,语音交互方法的各个步骤均依赖于声学语义大模型1000的各个模块。即,步骤021及其子步骤0211和子步骤0212基于声学编码模块100实现,步骤022及其子步骤0221基于字符转写模块200实现,步骤023及其子步骤0231、子步骤0232和子步骤0233基于知识检索模块300实现,步骤024、步骤024的子步骤0241和子步骤0242、及子步骤0242的次步骤02421、次步骤02422和次步骤02423基于大语言模型模块400实现。其中,关于语音交互方法的方法步骤中涉及的技术术语和解释说明,请参考对声学语义大模型各个模块的详细描述,此处不再赘述。

[0160] 综上,本申请实施方式提供的语音交互方法,基于上述的声学语义大模型1000实现,能够迅速且准确地对用户语音请求进行处理,增强用户体验。

[0161] 本申请还提供了一种计算机可读存储介质,其上存储有计算机程序。当计算机程序处理器执行的情况下实现如上述的车辆控制方法的步骤。

[0162] 可以理解,计算机程序包括计算机程序代码。计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。计算机可读存储介质可以包括:能够携带计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、以及

软件分发介质等。

[0163] 在本说明书的描述中,参考术语“具体地”、“进一步地”、“特别地”、“可以理解地”等的描述意指结合实施方式或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施方式或示例中。在本说明书中,对上述术语的示意性表述不预定指的是相同的实施方式或示例。而且,描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施方式或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0164] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本申请的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本申请的实施例所属技术领域的技术人员所理解。

[0165] 尽管上面已经示出和描述了本申请的实施方式,可以理解的是,上述实施方式是示例性的,不能理解为对本申请的限制,本领域的普通技术人员在本申请的范围内可以对上述实施方式的变化、修改、替换和变型。

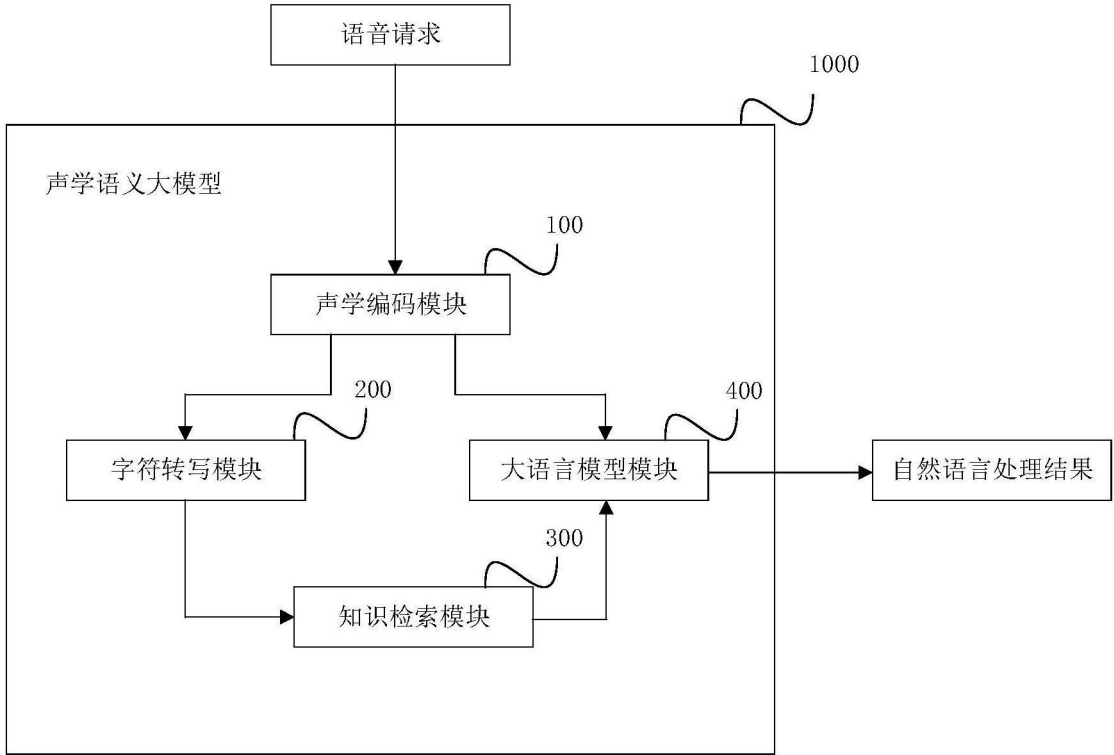


图1

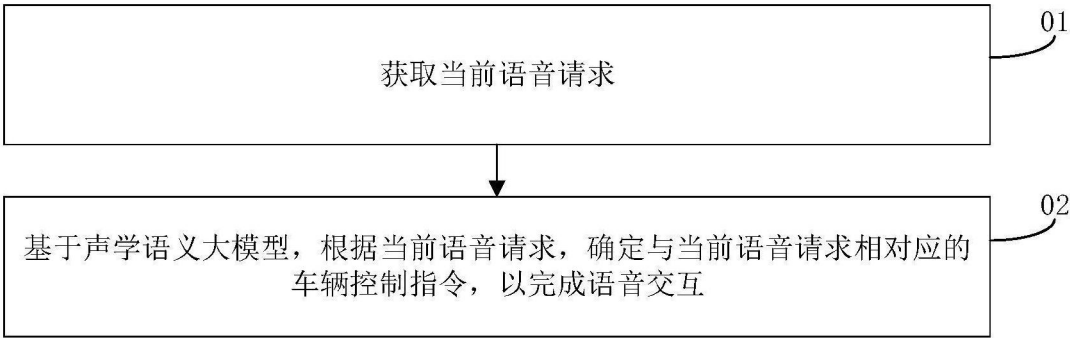


图2

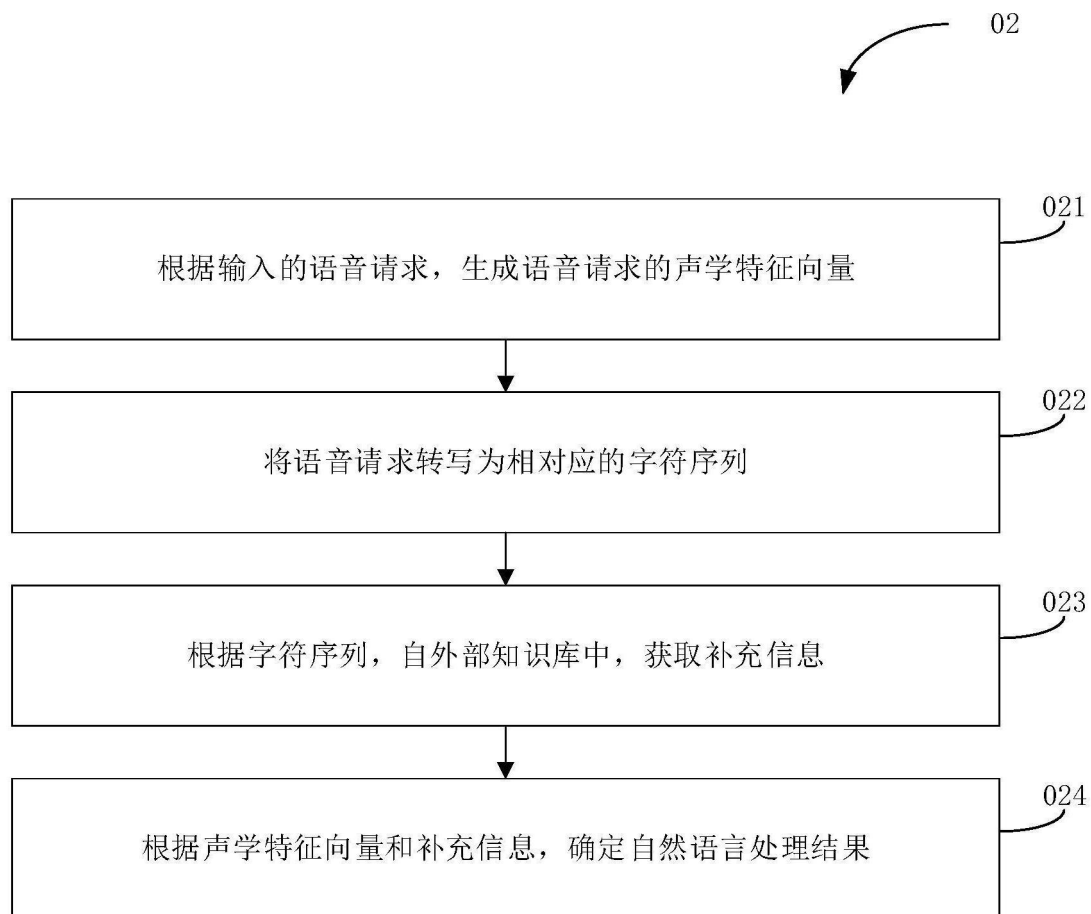


图3

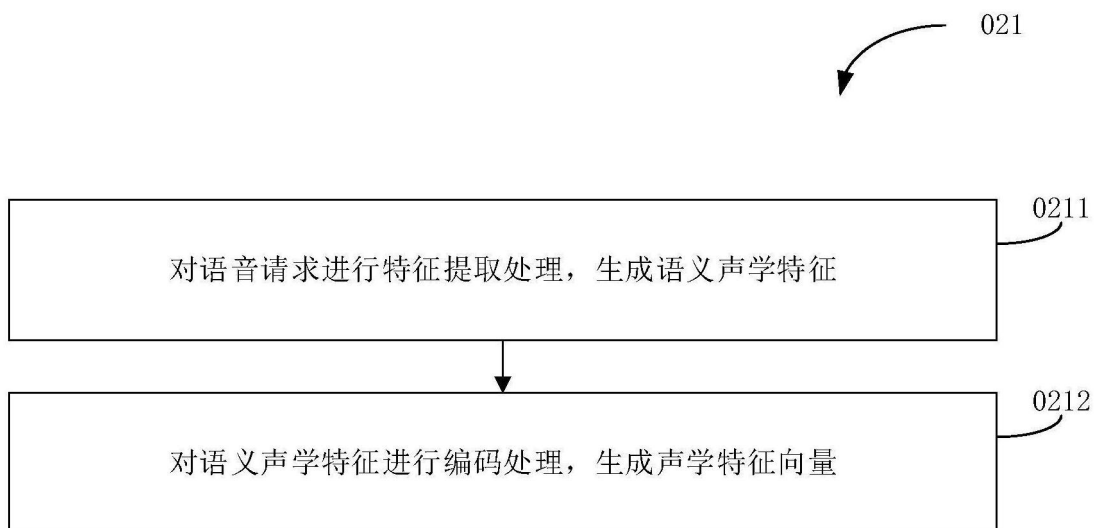


图4

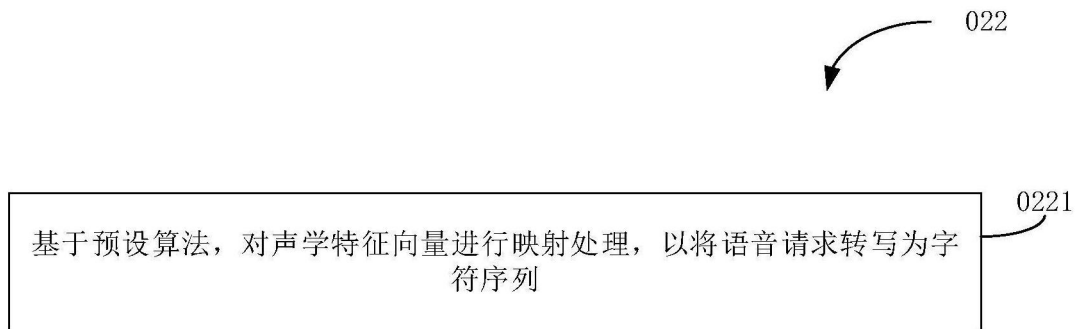


图5

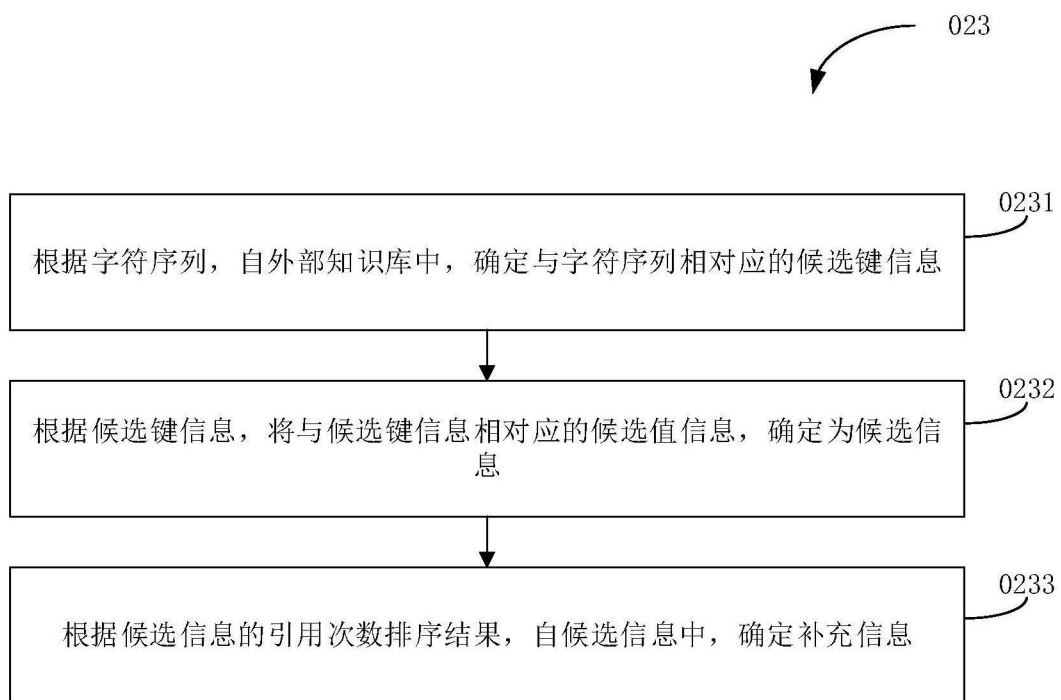


图6

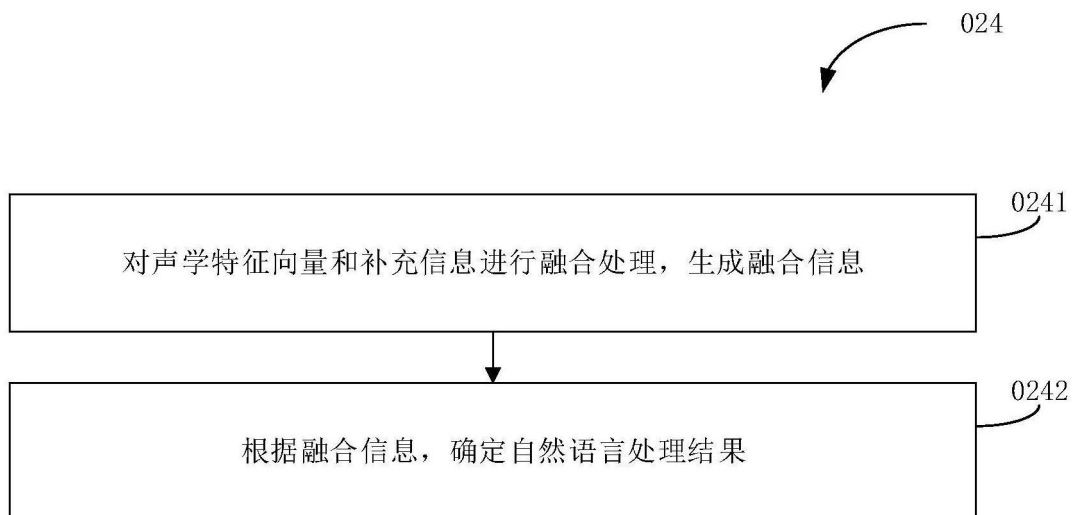


图7

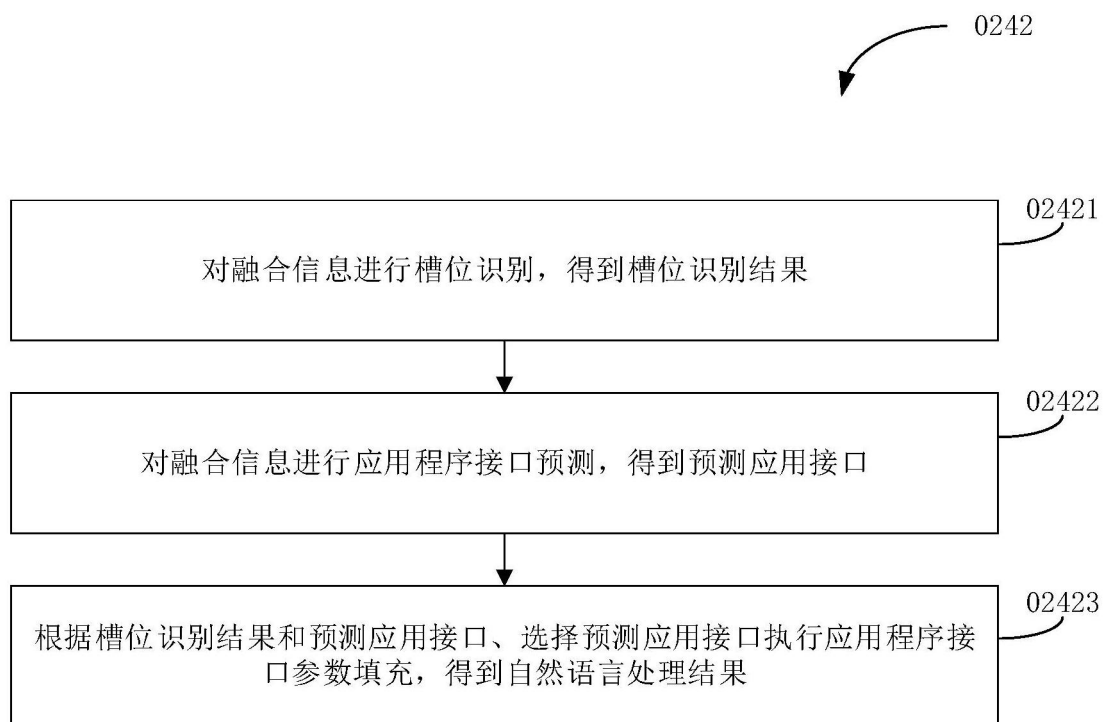


图8