(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117975964 A
(43)申请公布日 2024.05.03

(21)申请号 202410331756.8

(22)申请日 2024.03.22

(71)申请人 联众智慧科技股份有限公司
地址 310000 浙江省杭州市滨江区滨安路1197号2幢4166室

(72)发明人 张楚俊 潘豪格 金迪

(74)专利代理机构 北京恒泰铭睿知识产权代理有限公司 11642
专利代理师 郭建明

(51)Int.Cl.
G10L 15/22(2006.01)
G10L 15/24(2013.01)
G10L 15/18(2013.01)

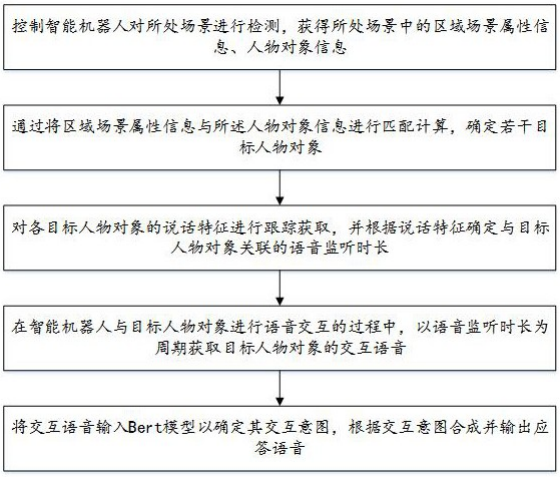
权利要求书2页 说明书9页 附图1页

(54)发明名称

一种基于Bert模型的智能机器人语音交互方法及其系统

(57)摘要

本发明属于智能机器人技术领域。提供了一种基于Bert模型的智能机器人语音交互方法及其系统。其中,所述方法包括:控制智能机器人对所处场景进行检测,获得所处场景中的区域场景属性信息、人物对象信息;通过将区域场景属性信息与人物对象信息进行匹配计算,确定若干目标人物对象;对各目标人物对象的说话特征进行跟踪获取,并根据说话特征确定与目标人物对象关联的语音监听时长;在智能机器人与目标人物对象进行语音交互的过程中,以语音监听时长为周期获取目标人物对象的交互语音;将交互语音输入Bert模型以确定其交互意图,根据交互意图合成并输出应答语音。本发明通过分析人物的交互语言的中断特点,提升了语音交互效果。



CN 117975964 A

1.一种基于Bert模型的智能机器人语音交互方法,其特征在于:所述方法包括如下步骤:

控制智能机器人对所处场景进行检测,获得所处场景中的区域场景属性信息、人物对象信息;

通过将所述区域场景属性信息与所述人物对象信息进行匹配计算,确定若干目标人物对象;

对各所述目标人物对象的说话特征进行跟踪获取,并根据所述说话特征确定与所述目标人物对象关联的语音监听时长;

在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的交互语音;

将所述交互语音输入Bert模型以确定其交互意图,根据所述交互意图合成并输出应答语音。

2.根据权利要求1所述的一种基于Bert模型的智能机器人语音交互方法,其特征在于:所述控制智能机器人对所处场景进行检测,获得所处场景中的区域场景属性信息、人物对象信息,包括:

对所处场景进行全景图像摄取,从摄取的全景图像中确定得出若干标识信息以及人物对象信息;

对所述标识信息进行语义分析以确定出与各个所述标识信息对应的区域场景属性信息。

3.根据权利要求2所述的一种基于Bert模型的智能机器人语音交互方法,其特征在于:所述通过将所述区域场景属性信息与所述人物对象信息进行匹配计算,确定若干目标人物对象,包括:

根据所述人物对象信息对各人物对象进行轨迹跟踪,根据所述轨迹与各区域场景的交叉关系确定出对应的各中间区域场景;其中,所述交叉关系包括交叉点信息和交叉时长信息;

根据与所述中间区域场景对应的所述区域场景属性信息分析得出各所述中间区域场景的行为终止概率;其中,所述行为终止概率指的是人物对象从中间区域场景直接离开所处场景的概率;

将所述行为终止概率高于指定阈值的所述中间区域场景确定所述目标区域场景,将位于所述目标区域场景外的人物对象确定为目标人物对象。

4.根据权利要求1所述的一种基于Bert模型的智能机器人语音交互方法,其特征在于:所述对各所述目标人物对象的说话特征进行跟踪获取,并根据所述说话特征确定与所述目标人物对象关联的语音监听时长,包括:

跟踪获取所述目标人物对象在与智能机器人进行语音交互之前的说话语音,将所述说话语音转换为语音文本;所述语音文本包含文本内容及嵌入其中的多个标点符号,以及与各标点符号对应的时刻;

基于标点符号将所述语音文本划分为多个子文本,根据所述子文本的起讫时刻确定所述子文本的第一说话时长;

根据各所述第一说话时长拟合得出第一概率分布曲线,将所述第一概率分布曲线与预

存的各第二概率分布曲线进行匹配计算,获得命中的所述第二概率分布曲线;

根据命中的所述第二概率分布曲线的峰值点确定得出第二说话时长,将所述第二说话时长确定为与所述目标人物对象关联的所述语音监听时长。

5.根据权利要求4所述的一种基于Bert模型的智能机器人语音交互方法,其特征在于:所述在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的交互语音,包括:

在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的第一交互语音;

在获取所述第一交互语音的过程中,同步获取所述目标人物对象的眼部动特征数据;

基于所述眼部动特征数据确定监听延长指数,将所述监听延长指数与所述语音监听时长相乘,获得新的所述语音监听时长;

以新的所述语音监听时长为周期获取所述目标人物对象的交互语音。

6.根据权利要求5所述的一种基于Bert模型的智能机器人语音交互方法,其特征在于:所述眼部动特征数据包括眼睛视线方向、眼周肌肉运动数据;则所述基于所述眼部动特征数据确定监听延长指数,包括:

调用AI分析模块对所述眼睛视线方向和所述眼周肌肉运动数据进行同步处理,获得表达障碍评估概率值,根据所述表达障碍评估概率值确定得出所述监听延长指数。

7.一种基于Bert模型的智能机器人语音交互系统,应用于远程智能监控终端,所述系统包括语音监听模组、摄像模组、处理模组、存储模组,所述处理模组分别与所述存储模组、所述语音监听模组、所述摄像模组电连接;

所述语音监听模组,用于监听目标人物对象的语音,并传输给所述处理模组;

所述摄像模组,用于获取智能机器人所处场景的图像及目标人物对象的图像,并传输给所述处理模组;

所述存储模组,用于存储计算机程序;

其特征在于:所述处理模组,用于调取并执行所述存储模组中的计算机程序,以执行如权利要求1-7任一所述的方法,以确定目标人物对象的交互意图,并根据所述交互意图合成并输出应答语音。

8.一种电子设备,包括:至少一个处理器、存储器以及存储在所述存储器中并可在所述至少一个处理器上运行的计算机程序,其特征在于:所述处理器执行所述计算机程序以实现如权利要求1-7任一所述的方法。

9.一种计算机存储介质,所述计算机存储介质存储有计算机程序,其特征在于:所述计算机程序被处理器执行以实现如权利要求1-7任一所述的方法。

10.一种计算机程序产品,包括存储于非暂时性计算机可读介质上的计算机程序,其特征在于:所述计算机程序被处理器执行时实现如权利要求1-6任一项所述的方法。

一种基于Bert模型的智能机器人语音交互方法及其系统

技术领域

[0001] 本发明涉及智能机器人技术领域,具体而言,涉及一种基于Bert模型的智能机器人语音交互方法及其系统。

背景技术

[0002] BERT(Bidirectional Encoder Representation from Transformers)是一个预训练的语言表征模型。它强调了不再像以往一样采用传统的单向语言模型或者把两个单向语言模型进行浅层拼接的方法进行预训练,而是采用新的masked language model(MLM),以致能生成深度的双向语言表征。由于Bert模型强大的语言表征能力,越来越多的机器人开始使用Bert模型实现与人的语音交互。

[0003] 专利文献1(CN117354591A)公开了一种语音交互式有线电视视频推荐方法,包括以下步骤:S1:获取待识别的音频信号,对音频信号进行预处理,提取预处理后的声学特征,基于声学特征构建声学模型,并对其进行训练输出文本信息;S2:基于输出的文本信息,利用BERT模型进行意图分析,通过对BERT模型进行预训练、调整,识别文本信息的意图;S3:基于文本意图识别、数据画像的分析,推荐与文本信息的意图相匹配的电视视频。

[0004] 专利文献2(CN115547313A)公开了一种基于驾驶员语音控制行驶车辆急停的方法,包括如下步骤:获取驾驶车辆的驾驶员的语音信息;采用BERT-L模型对驾驶员的语音信息向量化;将向量化的词向量群与车载终端预存储的向量词库进行匹配,以确认是否存在停车指令;若匹配成功,则获得停车指令,且检测车辆的车速达到预设速度时,执行停车指令,控制行驶车辆紧急停车。

[0005] 专利文献3(CN116303920A)公开了一种对话系统中的小样本商品规格信息识别和提取方法,包括以下步骤:S100:通过对话系统获取包括有商品规格信息的多个原始样本语料;S200:对所述原始样本语料进行数据扩充和增强,生成扩充增强样本语料,并进行预标注;S300:分别提取所述扩充增强样本语料的语音稀疏特征,每个token在预训练bert特征向量的预训练稠密特征,以及所述token对应的查找表稀疏特征;S400:将所述语音稀疏特征、预训练稠密特征、查找表稀疏特征进行特征融合,得到NER模型;S500:对所述NER模型训练得到烟草规格信息识别模型,通过所述烟草规格信息识别模型对烟草信息进行识别。

[0006] 可见,已经有较多的现有技术将Bert模型应用于语音交互领域中,但是现有的语音交互仅关注于交互意图的分析及应答,较少考虑交互人员的语言表达特点,尤其是交互语言的中断特点,导致语音交互效果较差。

发明内容

[0007] 对此,本发明提供了一种基于Bert模型的智能机器人语音交互方法、系统、电子设备及计算机存储介质,以解决上述技术问题。

[0008] 本发明公开了一种基于Bert模型的智能机器人语音交互方法,所述方法包括如下步骤:

控制智能机器人对所处场景进行检测,获得所处场景中的区域场景属性信息、人物对象信息;

通过将所述区域场景属性信息与所述人物对象信息进行匹配计算,确定若干目标人物对象;

对各所述目标人物对象的说话特征进行跟踪获取,并根据所述说话特征确定与所述目标人物对象关联的语音监听时长;

在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的交互语音;

将所述交互语音输入Bert模型以确定其交互意图,根据所述交互意图合成并输出应答语音。

[0009] 在一些实施例中,所述控制智能机器人对所处场景进行检测,获得所处场景中的区域场景属性信息、人物对象信息,包括:

对所处场景进行全景图像摄取,从摄取的全景图像中确定得出若干标识信息以及人物对象信息;

对所述标识信息进行语义分析以确定出与各个所述标识信息对应的区域场景属性信息。

[0010] 在一些实施例中,所述通过将所述区域场景属性信息与所述人物对象信息进行匹配计算,确定若干目标人物对象,包括:

根据所述人物对象信息对各人物对象进行轨迹跟踪,根据所述轨迹与各区域场景的交叉关系确定出对应的各中间区域场景;其中,所述交叉关系包括交叉点信息和交叉时长信息;

根据与所述中间区域场景对应的所述区域场景属性信息分析得出各所述中间区域场景的行为终止概率;其中,所述行为终止概率指的是人物对象从中间区域场景直接离开所处场景的概率;

将所述行为终止概率高于指定阈值的所述中间区域场景确定所述目标区域场景,将位于所述目标区域场景外的人物对象确定为目标人物对象。

[0011] 在一些实施例中,所述对各所述目标人物对象的说话特征进行跟踪获取,并根据所述说话特征确定与所述目标人物对象关联的语音监听时长,包括:

跟踪获取所述目标人物对象在与智能机器人进行语音交互之前的说话语音,将所述说话语音转换为语音文本;所述语音文本包含文本内容及嵌入其中的多个标点符号,以及与各标点符号对应的时刻;

基于标点符号将所述语音文本划分为多个子文本,根据所述子文本的起讫时刻确定所述子文本的第一说话时长;

根据各所述第一说话时长拟合得出第一概率分布曲线,将所述第一概率分布曲线与预存的各第二概率分布曲线进行匹配计算,获得命中的所述第二概率分布曲线;

根据命中的所述第二概率分布曲线的峰值点确定得出第二说话时长,将所述第二说话时长确定为与所述目标人物对象关联的所述语音监听时长。

[0012] 在一些实施例中,所述在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的交互语音,包括:

在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的第一交互语音;

在获取所述第一交互语音的过程中,同步获取所述目标人物对象的眼部动特征数据;

基于所述眼部动特征数据确定监听延长指数,将所述监听延长指数与所述语音监听时长相乘,获得新的所述语音监听时长;

以新的所述语音监听时长为周期获取所述目标人物对象的交互语音。

[0013] 在一些实施例中,所述眼部动特征数据包括眼睛视线方向、眼周肌肉运动数据;则所述基于所述眼部动特征数据确定监听延长指数,包括:

调用AI分析模块对所述眼睛视线方向和所述眼周肌肉运动数据进行同步处理,获得表达障碍评估概率值,根据所述表达障碍评估概率值确定得出所述监听延长指数。

[0014] 本发明还公开了一种基于Bert模型的智能机器人语音交互系统,应用于远程智能监控终端,所述系统包括语音监听模组、摄像模组、处理模组、存储模组,所述处理模组分别与所述存储模组、所述语音监听模组、所述摄像模组电连接;

所述语音监听模组,用于监听目标人物对象的语音,并传输给所述处理模组;

所述摄像模组,用于获取智能机器人所处场景的图像及目标人物对象的图像,并传输给所述处理模组;

所述存储模组,用于存储计算机程序;

所述处理模组,用于调取并执行所述存储模组中的计算机程序,以执行如前任一所述的方法,以确定目标人物对象的交互意图,并根据所述交互意图合成并输出应答语音。

[0015] 本发明还公开了一种电子设备,包括:至少一个处理器、存储器以及存储在所述存储器中并可在所述至少一个处理器上运行的计算机程序,所述处理器执行所述计算机程序以实现如前述实施例所述的方法。

[0016] 本发明还公开了一种计算机存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行以实现如前任一所述的方法。

[0017] 本发明还公开了一种计算机程序产品,当计算机程序产品在终端上运行时,使得终端执行时以实现如前任一所述的方法。

[0018] 本发明的有益效果在于:

本发明中的智能机器人可以事先对所处场景内的人物对象进行综合分析,以确定存在与其进行语音交互概率的目标人物对象,然后对这些目标人物对象的说话特征进行跟踪获取,从而可以确定出针对该目标人物对象的语音监听时长,智能机器人便可以按照该语音监听时长来监听目标人物对象在与智能机器人进行语音交互过程中的单次语音交互内容,从而在确保目标人物对象语音交互内容完整的基础上提升语音应答效率。

附图说明

[0019] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本发明的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其它相关的附图。

[0020] 图1是本发明实施例公开的一种基于Bert模型的智能机器人语音交互方法的流程示意图；

图2是本发明实施例公开的一种基于Bert模型的智能机器人语音交互系统的结构示意图。

具体实施方式

[0021] 以下由特定的具体实施例说明本申请的实施方式，熟悉此技术的人士可由本说明书所揭露的内容轻易地了解本申请的其他优点及功效，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0022] 此外，下面所描述的本申请不同实施方式中所涉及的技术特征只要彼此之间未构成冲突就可以相互结合。

[0023] 如图1所示，本发明实施例公开了一种基于Bert模型的智能机器人语音交互方法，所述方法包括如下步骤：

控制智能机器人对所处场景进行检测，获得所处场景中的区域场景属性信息、人物对象信息；

通过将所述区域场景属性信息与所述人物对象信息进行匹配计算，确定若干目标人物对象；

对各所述目标人物对象的说话特征进行跟踪获取，并根据所述说话特征确定与所述目标人物对象关联的语音监听时长；

在智能机器人与所述目标人物对象进行语音交互的过程中，以所述语音监听时长为周期获取所述目标人物对象的交互语音；

将所述交互语音输入Bert模型以确定其交互意图，根据所述交互意图合成并输出应答语音。

[0024] 在本发明实施例中，不同人员进行语音交互时的语言表达特点是有较大差别的，主要体现在对于单个交互意图所采取的语言长度的不同。例如，在表达对某业务的咨询时，A用户思维清晰、语言表达能力强，其语音交互过程为“请问接待大集团客户的VIP室在哪边?”；而B用户思维较为模糊、思维清晰、语言表达能力差，其语音交互过程为“请问VIP室在哪边?…接待大集团客户的VIP室”。由于B用户在语音交互过程中存在较长时间的语音中断，智能机器人容易误判其单个交互意图已经输出完成，这样容易基于“请问VIP室在哪边?”输出对应的应答语音，导致交互准确性不足。

[0025] 本发明中的智能机器人可以事先对所处场景内的人物对象进行综合分析，以确定存在与其进行语音交互概率的目标人物对象，然后对这些目标人物对象的说话特征进行跟踪获取，从而可以确定出针对该目标人物对象的语音监听时长，智能机器人便可以按照该语音监听时长来监听目标人物对象在与智能机器人进行语音交互过程中的单次语音交互内容，从而在确保目标人物对象语音交互内容完整的基础上提升语音应答效率。

[0026] 在一些实施例中，所述控制智能机器人对所处场景进行检测，获得所处场景中的区域场景属性信息、人物对象信息，包括：

对所处场景进行全景图像摄取，从摄取的全景图像中确定得出若干标识信息以及

人物对象信息；

对所述标识信息进行语义分析以确定出与各个所述标识信息对应的区域场景属性信息。

[0027] 在本发明实施例中，智能机器人配备有摄像头，通过控制智能机器人进行旋转便可以完成对所处场景的全景图像的摄取，摄取的全景图像覆盖了所处场景内的全部区域。然后，再按照特定的提取规则从全景图像中识别并提取出各个标识信息以及人物对象信息。标识信息可以是所处场景内的“出口标识”、“入口标识”、“普通柜台”、“VIP柜台”、“等候区”等，通过对标识信息进行语义分析便可以确定这些标识信息所对应区域的区域场景属性信息。当然，由于所处场景的不同，上述标识信息也会存在不同。

[0028] 其中，也可以使用前述的Bert模型来分析上述标识信息的词义/语义，从而确定出对应区域的区域场景属性信息。

[0029] 在一些实施例中，所述通过将所述区域场景属性信息与所述人物对象信息进行匹配计算，确定若干目标人物对象，包括：

根据所述人物对象信息对各人物对象进行轨迹跟踪，根据所述轨迹与各区域场景的交叉关系确定出对应的各中间区域场景；其中，所述交叉关系包括交叉点信息和交叉时长信息；

根据与所述中间区域场景对应的所述区域场景属性信息分析得出各所述中间区域场景的行为终止概率；其中，所述行为终止概率指的是人物对象从中间区域场景直接离开所处场景的概率；

将所述行为终止概率高于指定阈值的所述中间区域场景确定所述目标区域场景，将位于所述目标区域场景外的人物对象确定为目标人物对象。

[0030] 在本发明实施例中，智能机器人可继续对所处场景内的人物对象进行跟踪以确定其轨迹，再计算轨迹与各区域场景的交叉关系，包括是否存在交叉点、轨迹在各区域场景内的交叉时长(即区域场景内的停留时长)，这样就可以确定出该人物对象在被跟踪过程中所进入的各个中间区域场景。接着，对各中间区域场景所对应的区域场景属性信息进行深入分析可以得出该人物对象从该中间区域场景直接离开所处场景的概率，该处的直接离开的含义是人物对象从该中间区域场景直接行至所处场景的出口处，且该驶离过程中，在其它中间区域场景的停留时长短于特定值(例如2s)，例如人物对象在VIP柜台办完业务之后径直走向出口而离开所处场景，中间虽然可能经过普通柜台区域但停留时长小于2s。在其概率高于指定阈值时即可将该中间区域场景认定为目标区域场景。由于目标区域场景以内的人物会直接离开所处场景，其几乎不存在与智能机器人进行语音交互的可能，无需对其说话特征进行分析，所以将位于目标区域场景以外的人物确定为目标人物对象。

[0031] 举例说明如下：某中间区域场景为“普通柜台”或“VIP柜台”，对其与“普通柜台”或“VIP柜台”对应的区域场景属性信息进行深入分析可知该中间区域场景为业务办理区域，位于该区域内的人物将会在完成业务办理之后直接从“出口”离开所处场景。显然，该人物不会存在与智能机器人进行语音交互的需求，或者语音交互的概率较低，无需将其作为目标人物对象以进行说话特征的分析。

[0032] 与此对应地，位于“等候区”的人物对象尚未完成业务办理，其行为终止概率低于指定阈值，位于“等候区”内的人物不会直接从“出口”离开所处场景，需要将其作为目标人

物对象以进行说话特征的分析。

[0033] 需要说明的是,有些场景内的业务可能需要多个柜台接力办理才可以完成,可以提前根据对应场景内的业务办理实际分派布设情况,而确定各中间区域场景的行为终止概率。

[0034] 在一些实施例中,所述对各所述目标人物对象的说话特征进行跟踪获取,并根据所述说话特征确定与所述目标人物对象关联的语音监听时长,包括:

跟踪获取所述目标人物对象在与智能机器人进行语音交互之前的说话语音,将所述说话语音转换为语音文本;所述语音文本包含文本内容及嵌入其中的多个标点符号,以及与各标点符号对应的时刻;

基于标点符号将所述语音文本划分为多个子文本,根据所述子文本的起讫时刻确定所述子文本的第一说话时长;

根据各所述第一说话时长拟合得出第一概率分布曲线,将所述第一概率分布曲线与预存的各第二概率分布曲线进行匹配计算,获得命中的所述第二概率分布曲线;

根据命中的所述第二概率分布曲线的峰值点确定得出第二说话时长,将所述第二说话时长确定为与所述目标人物对象关联的所述语音监听时长。

[0035] 在本发明实施例中,受思维敏捷度、口齿生理特性等因素的影响,人在说话时的节奏是存在较大差异的。所以本发明通过对前述确定出的目标人物对象的说话语音进行跟踪,从中提取出目标人物对象的说话特征,再依据该说话特征中的单次语音表达时长特征来确定语音监听时长,语音监听时长指的是智能机器人在收听目标人物对象的交互语音时的等待时长,即在该语音监听时长内智能机器人不进行交互语音的识别,而是等目标人物对象完成单次语义的输入之后再进行语义分析。

[0036] 具体实施时,先利用现有的语义识别模型(例如Bert模型)对在其它区域内正在进行说话的目标人物对象进行监听,并对目标人物对象与智能机器人进行语音交互之前的全部说话语音收集后再整体进行语义分析,从而获得对应的语音文本。该语音文本中包含了具有独立语义的各个语句,各语句间通过标点符号进行连接,而各标点符号对应着该具有独立语义的语句在说出时的起讫时刻。以语音文本中的标点符号为分隔点将该语音文本分为多个子文本,从而可以确定各个子文本对应的第一说话时长。对全部第一说话时长进行曲线拟合,从而可以得到一个第一概率分布曲线,同时数据库中预存了多个第二概率分布曲线,第一概率分布曲线和第二概率分布曲线服从正态分布、指数分布等,实际测试过程中较多符合类正态分布,尤其是在短时对话过程中的说话语音更多属于类正态分布,这是源于短时对话过程整体来说包含一个主题的对话的内容。通过对第一概率分布曲线和第二概率分布曲线进行相似度计算便可以命中一个第二概率分布曲线,基于第二概率分布曲线的峰值点(即时长)确定得出第二说话时长,第二说话时长即表征了符合该目标人物对象的说话中断特点(可由各第一说话时长拟合得出的上述第一概率分布曲线表示)的单次输出具有独立语义的语音的最大时长、平均时长等,将第二说话时长作为语音监听时长。于是,智能机器人按照该语音监听时长来监听目标人物对象单次输出具有独立语义的语音,可以确保该具有独立语义的语音能够被语音监听时长覆盖,又不会将语音监听时长设置的过长,设置的过长时会导致智能机器人无效监听等待时长过长,这样就不利于后续的语音应答。

[0037] 需要说明的是,智能机器人可以通过自身配置的语音监听模组来监听目标人物对

象的说话语音,也可以通过分布式布设于所处场景中的麦克拾音器来获取与智能机器人处于较远距离的目标人物对象的说话语音。

[0038] 在一些实施例中,所述在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的交互语音,包括:

在智能机器人与所述目标人物对象进行语音交互的过程中,以所述语音监听时长为周期获取所述目标人物对象的第一交互语音;

在获取所述第一交互语音的过程中,同步获取所述目标人物对象的眼部动特征数据;

基于所述眼部动特征数据确定监听延长指数,将所述监听延长指数与所述语音监听时长相乘,获得新的所述语音监听时长;

以新的所述语音监听时长为周期获取所述目标人物对象的交互语音。

[0039] 在本发明实施例中,通过前述方式确定出的语音监听时长可以在理论范围内匹配目标人物对象的说话中断特点,但由于前述确定出的第一概率分布曲线实际上仅是基于目标人物对象的短时说话语音,其可信度实际上并不能完全保证。所以,本发明设置在智能机器人与目标人物对象进行语音对话的过程中,先按照前述确定得出的语音监听时长为周期获取目标人物对象的交互语音,在此过程中同步获取该目标人物对象的眼部动特征数据,根据眼部动特征数据来决策是否将语音监听时长适当延长。

[0040] 在一些实施例中,所述眼部动特征数据包括眼睛视线方向、眼周肌肉运动数据;则所述基于所述眼部动特征数据确定监听延长指数,包括:

调用AI分析模块对所述眼睛视线方向和所述眼周肌肉运动数据进行同步处理,获得表达障碍评估概率值,根据所述表达障碍评估概率值确定得出所述监听延长指数。

[0041] 在本发明实施例中,人会因为口吃、思维没跟上(例如暂未想到合适的表达词汇)等原因而出现表达障碍的情况,从而会出现说话语句异常中断的情况,如果不对前述确定出的语音监听时长进行适当延长,则可能会出现目标人物对象在克服了表达障碍之后的说话语音不能被划入本次的监听周期,从而导致本监听周期中的说话语音实际上是不完整的,最终会导致提取出的交互意图失真,而且由于目标人物对象在克服了表达障碍之后的说话语音被划入下一监听周期,还可能会导致下一监听周期的交互意图识别结果的失真。

[0042] 所以,本发明基于目标人物对象在对智能机器人输出说话语音时的眼部动特征数据来分析其是否出现了表达障碍,眼部动特征数据主要包括眼睛视线方向、眼周肌肉运动数据。其中,眼睛视线方向包括直视(对准摄像头或与摄像头偏离指定角度内,即目标人物对象在正视智能机器人)、斜向左/右上、斜向左/右下等,微表情相关研究结果表明,在人的视线位于斜向左上、右上时,一般是在进行快速的思考,所以可以基于该视线特征来分析目标人物对象是否因为思维没跟上(例如暂未想到合适的表达词汇)的原因而出现表达障碍。眼周肌肉运动数据则指的是目标人物对象出现口吃时的眼周肌肉异常动作特征,因为当舌头的肌肉发生运动异常时会导致控制眼睛的肌肉也发生运动异常。同时,本发明优选建立用于分析上述眼部动特征数据的AI分析模块,使用其分析计算出目标人物对象的表达障碍评估概率值,然后再按照例如正比例函数来确定出适宜的监听延长指数,使用监听延长指数可以实现将前述确定出的语音监听时长进行适当延长,新的语音监听时长则会大概率覆盖到表达障碍及后续的说话语音的时段,这样就会最终监听到用于分析具有独立语义的语

音的全部说话语音。

[0043] 举例说明如下：

智能机器人首先以语音监听时长为5s的周期获取目标人物对象的第一交互语音，在此过程中对目标人物对象的眼睛视线方向、眼周肌肉运动数据进行检测分析，在计算出的表达障碍评估概率值超出了正常范围时，则可按照表达障碍评估概率值和对应的正比例函数确定出监听延长指数例如为1.5，再将其与原来的语音监听时长相乘便得到，5*1.5=7.5s。而7.5s的语音监听时长就可以覆盖目标人物对象本次出现的表达障碍及剩余说话语音的所需时间。将在该周期内监听到的交互语音输入至Bert模型便可以确定出完整且准确的交互意图，有利于后续输出匹配的应答语音。

[0044] 需要说明的是，本发明中的AI分析模块是预先建立的，在需要使用时由智能机器人进行调用，AI分析模块可以存储于智能机器人或网络服务器中。而且，AI分析模块已经经过了足够的预训练，能够基于眼睛视线方向、眼周肌肉运动数据分析出准确的表达障碍评估概率值。

[0045] 如图2所示，本发明实施例还公开了一种基于Bert模型的智能机器人语音交互系统，应用于远程智能监控终端，所述系统包括语音监听模组、摄像模组、处理模组、存储模组，所述处理模组分别与所述存储模组、所述语音监听模组、所述摄像模组电连接；

所述语音监听模组，用于监听目标人物对象的语音，并传输给所述处理模组；

所述摄像模组，用于获取智能机器人所处场景的图像及目标人物对象的图像，并传输给所述处理模组；

所述存储模组，用于存储计算机程序；

所述处理模组，用于调取并执行所述存储模组中的计算机程序，以执行如前任一所述的方法，以确定目标人物对象的交互意图，并根据所述交互意图合成并输出应答语音。

[0046] 本发明实施例还公开了一种电子设备，包括：至少一个处理器、存储器以及存储在所述存储器中并可在所述至少一个处理器上运行的计算机程序，其特征在于：所述处理器执行所述计算机程序以实现如前述实施例所述的方法。

[0047] 本发明实施例还公开了一种计算机存储介质，所述计算机存储介质存储有计算机程序，其特征在于：所述计算机程序被处理器执行以实现如前述实施例所述的方法。

[0048] 本发明实施例还公开了一种计算机程序产品，当计算机程序产品在终端上运行时，使得终端执行时以实现如前述实施例所述的方法。

[0049] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0050] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0051] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0052] 以上所述,仅为本发明的较佳实施例而已,并非用于限定本发明的保护范围。

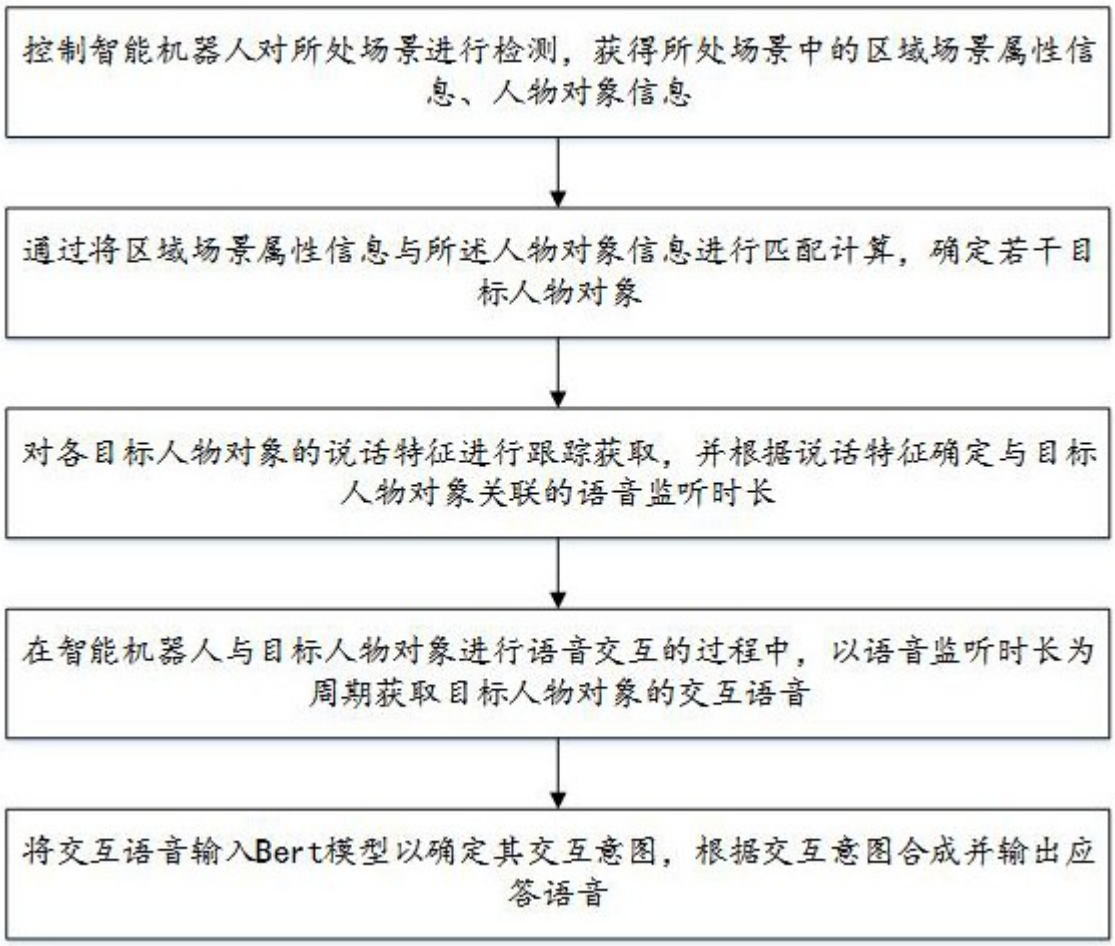

控制智能机器人对所处场景进行检测，获得所处场景中的区域场景属性信息、人物对象信息
通过将区域场景属性信息与所述人物对象信息进行匹配计算，确定若干目标人物对象
对各目标人物对象的说话特征进行跟踪获取，并根据说话特征确定与目标人物对象关联的语音监听时长
在智能机器人与目标人物对象进行语音交互的过程中，以语音监听时长为周期获取目标人物对象的交互语音
将交互语音输入Bert模型以确定其交互意图，根据交互意图合成并输出应答语音

图 1

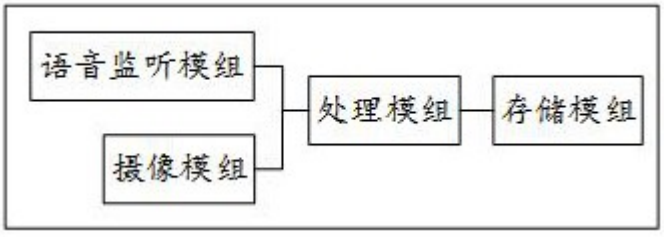

语音监听模组
摄像模组
处理模组
存储模组

图 2