



(12) 发明专利

(10) 授权公告号 CN 113808593 B

(45) 授权公告日 2025. 06. 03

(21) 申请号 202010552193.7

G10L 15/08 (2006.01)

(22) 申请日 2020.06.16

G10L 15/22 (2006.01)

(65) 同一申请的已公布的文献号

H04N 21/45 (2011.01)

申请公布号 CN 113808593 A

H04N 21/482 (2011.01)

H04N 21/239 (2011.01)

(43) 申请公布日 2021.12.17

(56) 对比文件

(73) 专利权人 阿里巴巴集团控股有限公司

CN 102722525 A, 2012.10.10

地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

CN 105392035 A, 2016.03.09

CN 107832434 A, 2018.03.23

(72) 发明人 郑梓豪 胡于响 姜飞俊 张帆

CN 110866099 A, 2020.03.06

CN 111145756 A, 2020.05.12

(74) 专利代理机构 北京睿派知识产权代理有限公司 11597

US 2015279365 A1, 2015.10.01

专利代理师 刘锋

审查员 樊冲

(51) Int. Cl.

G10L 15/30 (2013.01)

G10L 15/26 (2006.01)

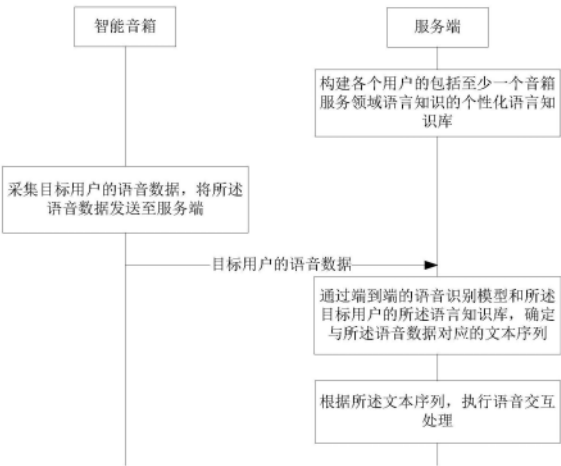
权利要求书11页 说明书29页 附图5页

(54) 发明名称

语音交互系统、相关方法、装置及设备

(57) 摘要

本申请公开了语音交互系统、相关方法、装置及设备。其中,所述系统通过智能音箱采集目标用户的语音数据,将所述语音数据发送至服务端;服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。采用这种处理方式,使得能够根据实时更新的知识库进行语音识别,因此可以有效兼顾较高的语音识别实时性和准确度。



1. 一种语音交互系统,其特征在于,包括:

智能音箱,用于采集目标用户的语音数据,将所述语音数据发送至服务端;

服务端,用于构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

2. 一种语音交互方法,其特征在于,包括:

构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;

针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;

根据所述文本序列,执行语音交互处理。

3. 根据权利要求2所述的方法,其特征在于,

所述根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征,包括:

通过所述语音识别模型包括的指示器评分模型,根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征。

4. 根据权利要求3所述的方法,其特征在于,

通过所述指示器评分模型,确定与所述第一文本特征对应的词与各个语言知识的相关度;

至少根据与所述相关度大于相关度阈值的词相关的语言知识,确定第二文本特征。

5. 根据权利要求2所述的方法,其特征在于,

所述方法还包括:

将语言知识的编码数据存储至所述语音识别模型包括的语言知识存储模块;

所述根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征,包括:

根据所述语言知识存储模块存储的所述编码数据和所述第一文本特征,确定第二文本特征。

6. 根据权利要求2所述的方法,其特征在于,还包括:

从训练数据集中学习得到所述语音识别模型;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息。

7.根据权利要求6所述的方法,其特征在于,所述训练数据中的个性化语言知识库,采用如下方式确定:

根据多个训练数据的文本序列标注信息,构建所述个性化语言知识库。

8.根据权利要求2所述的方法,其特征在于,

所述个性化语言知识库包括:长尾实体词,反语言模型的实体词,同音不同字的实体词,上下文中的实体词。

9.根据权利要求2所述的方法,其特征在于,

所述音箱服务领域包括:打电话服务领域;

所述打电话服务领域的语言知识包括:用户通讯录中的人名;

所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:

接收与所述用户对应的智能音箱发送的用户通讯录信息;

将用户通讯录中的人名作为所述用户的个性化语言知识。

10.根据权利要求2所述的方法,其特征在于,

所述音箱服务领域包括:问答服务领域;

所述问答服务领域的语言知识包括:上下文中的文本片段;

所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:

确定上下文文本序列;

将所述上下文文本序列中的文本片段作为所述用户的个性化语言知识。

11.根据权利要求2所述的方法,其特征在于,

所述音箱服务领域包括:多媒体播放服务领域;

所述多媒体播放服务领域的语言知识包括:曲目名;

所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:

确定用户的历史播放节目名;

将所述历史播放节目名作为所述用户的个性化语言知识。

12.根据权利要求2所述的方法,其特征在于,所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,采用如下方式的至少一个:

根据所述用户的购物数据,确定所述用户的个性化语言知识;

根据所述用户输入的文本信息,确定所述用户的个性化语言知识。

13.根据权利要求2所述的方法,其特征在于,还包括:

根据交互语音数据,更新所述用户的语言知识库。

14.一种语音交互方法,其特征在于,包括:

采集目标用户的语音数据;

将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语

言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

15.一种语音交互方法,其特征在于,包括:

确定用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;

针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;

根据所述文本序列,执行语音交互处理。

16.一种电视节目播放方法,其特征在于,包括:

确定用户的个性化节目播放语言知识库;

针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音指令数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音指令数据对应的第二文本特征;至少根据第二文本特征,确定所述目标节目名,其中,所述至少根据第二文本特征,确定所述目标节目名,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据融合后的文本特征确定目标节目名;

根据所述目标节目名,播放目标节目对象。

17.根据权利要求16所述的方法,其特征在于,还包括:

将所述用户的历史播放节目对象的节目名、演员名和/或导演名作为所述用户的个性化节目播放语言知识。

18.根据权利要求16所述的方法,其特征在于,所述根据所述目标节目名,播放目标节目对象,包括:

根据节目表,确定与所述目标节目名对应的电视频道和播放时间;

根据所述播放时间和所述电视频道,确定目标节目对象;

播放所述目标节目对象。

19.根据权利要求18所述的方法,其特征在于,

所述根据所述播放时间和所述电视频道,确定目标节目对象,包括:

显示与所述目标节目名对应的至少一个电视频道在多个时间播放的多个节目对象；
将用户指定的节目对象作为目标节目对象。

20. 根据权利要求18所述的方法,其特征在於,还包括:

若所述节目表不包括所述目标节目名,则确定与所述目标节目名相关的节目名;

显示相关节目名;

若用户指定播放相关节目对象,则播放相关节目对象。

21. 一种会议记录方法,其特征在於,包括:

构建会议领域的语言知识库;

采集会议语音数据;

通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

22. 根据权利要求21所述的方法,其特征在於,还包括:

确定与会议语音数据对应的会议领域。

23. 一种会议记录方法,其特征在於,包括:

构建各个领域的语言知识库;

针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;

通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

24. 一种语音识别模型构建方法,其特征在於,包括:

确定训练数据集;所述训练数据包括:语音数据、语言知识库、文本序列标注信息;

构建端到端的语音识别模型的网络结构;

从训练数据集中学习得到所述语音识别模型;其中,

所述语音识别模型包括语言知识编码器,用于对语言知识执行编码处理;

所述语音识别模型包括通用语言模型,用于确定与所述语音数据对应的第一文本特征;

所述语音识别模型,还用于根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述

至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

25. 根据权利要求24所述的方法,其特征在于,

所述模型包括:声音编码器、解码器和分类器。

26. 根据权利要求24所述的方法,其特征在于,

所述模型包括:声音编码器、解码器、语言模型、特征融合模块和分类器。

27. 根据权利要求24所述的方法,其特征在于,

所述模型包括:声音编码器、语言模型、指示器评分模型、特征融合模块和分类器。

28. 一种语音识别方法,其特征在于,包括:

构建用户的包括至少一个领域语言知识的个性化语言知识库;

针对终端设备采集的用户语音数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

29. 一种语音交互装置,其特征在于,包括:

知识库构建单元,用于构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;

语音识别单元,用于针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列,具体包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;

指令处理单元,用于根据所述文本序列,执行语音交互处理。

30. 一种电子设备,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码

数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;根据所述文本序列,执行语音交互处理。

31.一种语音交互装置,其特征在于,包括:

语音数据采集单元,用于采集目标用户的语音数据;

语音数据发送单元,用于将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;根据所述文本序列,执行语音交互处理。

32.一种智能音箱,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标用户的语音数据;将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;根据所述文本序列,执行语音交互处理。

33.一种电子设备,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标用户的语音数据;将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特

征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;根据所述文本序列,执行语音交互处理。

34.一种语音交互装置,其特征在于,包括:

知识库构建单元,用于确定用户的包括至少一个领域语言知识的个性化语言知识库;

语音识别单元,用于针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;

指令处理单元,用于根据所述文本序列,执行语音交互处理。

35.一种智能音箱,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列;根据所述文本序列,执行语音交互处理。

36.一种电子设备,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定用户的包括至少一个领域语言知识的个性化语言知识库;针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据

所述融合后的文本特征确定所述文本序列;根据所述文本序列,执行语音交互处理。

37. 一种电视节目播放装置,其特征在于,包括:

知识库确定单元,用于智能电视确定用户的个性化节目播放语言知识库;

节目识别单元,用于针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音指令数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音指令数据对应的第二文本特征;至少根据第二文本特征,确定所述目标节目名,其中,所述至少根据第二文本特征,确定所述目标节目名,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据融合后的文本特征确定目标节目名;

节目播放单元,用于根据所述目标节目名,播放目标节目对象。

38. 一种智能电视,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定用户的个性化节目播放语言知识库;针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音指令数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音指令数据对应的第二文本特征;至少根据第二文本特征,确定所述目标节目名,其中,所述至少根据第二文本特征,确定所述目标节目名,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据融合后的文本特征确定目标节目名;根据所述目标节目名,播放目标节目对象。

39. 一种电视节目播放装置,其特征在于,包括:

节目名识别单元,用于确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名,具体包括:确定用户的个性化节目播放语言知识库;针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音指令数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音指令数据对应的第二文本特征;至少根据第二文本特征,确定所述目标节目名,其中,所述至少根据第二文本特征,确定所述目标节目名,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据融合后的文本特征确定目标节目名;

节目对象确定单元,用于根据节目表,确定与所述目标节目名对应的目标节目对象;

播放单元,用于播放所述目标节目对象。

40. 一种智能电视,其特征在于,包括:

处理器和存储器；

所述存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名,具体包括:确定用户的个性化节目播放语言知识库;针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音指令数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音指令数据对应的第二文本特征;至少根据第二文本特征,确定所述目标节目名,其中,所述至少根据第二文本特征,确定所述目标节目名,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据融合后的文本特征确定目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象。

41. 一种会议记录装置,其特征在于,包括:

知识库构建单元,用于构建会议领域的语言知识库;

语音数据采集单元,用于采集会议语音数据;

语音转写单元,用于通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

42. 一种电子设备,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建会议领域的语言知识库;采集会议语音数据;通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

43. 一种会议记录装置,其特征在于,包括:

知识库构建单元,用于构建各个领域的语言知识库;

会议领域确定单元,用于针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;

语音转写单元,用于通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

44.一种电子设备,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建各个领域的语言知识库;针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

45.一种会议记录装置,其特征在于,包括:

语音数据采集单元,用于采集目标会议的语音数据;

语音数据发送单元,用于将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

46.一种电子设备,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标会议的语音数据;将所述语音数据发送至服务

端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

47.一种语音识别装置,其特征在于,包括:

知识库构建单元,用于构建用户的包括至少一个领域语言知识的个性化语言知识库;

模型预测单元,用于针对终端设备采集的用户语音数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

48.一种电子设备,其特征在于,包括:

处理器和存储器;

所述存储器,用于存储实现语音识别方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建用户的包括至少一个领域语言知识的个性化语言知识库;针对终端设备采集的用户语音数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音数据对应的文本序列,包括:通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;通过所述语音识别模型包括的通用语言模型,确定与所述语音数据对应的第一文本特征;根据语言知识的编码数据和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列;根据所述文本序列,执行语音交互处理,其中,所述至少根据第二文本特征,确定所述文本序列,包括:通过所述语音识别模型包括的特征融合模块,将所述第一文本特征和所述第二文本特征连接或相加,得到融合后的文本特征,根据所述融合后的文本特征确定所述文本序列。

语音交互系统、相关方法、装置及设备

技术领域

[0001] 本申请涉及数据处理技术领域,具体涉及语音交互系统、方法和装置,电视节目播放方法和装置,会议记录方法和装置,语音识别模型构建方法和装置,智能音箱,智能电视,以及电子设备。

背景技术

[0002] 智能音箱,是一个音箱升级的产物,是家庭消费者用语音进行上网的一个工具,比如点播歌曲、上网购物,或是了解天气预报,它也可以对智能家居设备进行控制,比如打开窗帘、设置冰箱温度、提前让热水器升温等。

[0003] 用户与智能音箱之间主要通过语音方式进行交互。用户向智能音箱下达语音指令,智能音箱通过自动语音识别(Automatic Speech Recognition,ASR)技术识别用户指令,并执行指令。在音箱类对话式交互系统中,存在千千万万的实体词,一方面长尾实体太多(如有声书名《宦妃权倾天下》),语言模型很难全部记忆,并且常常存在反语言模型的情况(如音乐歌名“情人结”,用户通讯录“梓豪vs子豪”),这类实体对ASR来说是个挑战,包括:a)语言模型训练样本有限,很难保证充分的覆盖;b)创作类型的实体,为求标新立异,常常“反语言模型”;c)语言模型解码空间相对开放,范围广。为了解决这个问题,目前应用于智能音箱的典型的语音识别系统是,首先为各个音箱技能构建其专用的子语言模型,子语言模型可涵盖相应技能涉及的长尾实体、反语言模型的实体、同音不同字的实体等;然后针对各个音箱技能的语音数据,通过通用的语言模型、及音箱技能专有的子语言模型,识别该音箱技能的语音数据。

[0004] 然而,在实现本发明过程中,发明人发现该技术方案至少存在如下问题:1) 当一个音箱技能出现新的语言知识时,相应的子语言模型在更新语言知识时,需要重新训练,会花费较多时间,由此导致无法兼顾较高的语音识别实时性和准确度;2) 子语言模型存在句式依赖的问题,如音箱的打电话技能包括“打电话给郑梓豪”这个句式,这样就会使得一个音箱技能的个性化语言知识(如“梓豪”)无法应用到其它音箱技能的语音识别中,从而影响语音识别准确度;3) 现有技术中的语音识别模型为非端到端的模型,该模型存在的误差传播问题,因而影响语音识别准确度;4) 智能音箱作为对话式交互系统,上下文的信息也十分重要,如最近很火的「你想我猜」技能中,用户正在猜动物,天猫精灵问“请问它有角吗?”,用户答复到“you jiao”;如果ASR不能准确的理解上文,则可能会将用户答复的语音“you jiao”识别为“有脚”这个错误的结果,游戏便难以顺畅的进行下去,而现有的语音识别模型无法根据对话内容实时更新子语言模型,因而影响语音识别准确度。综上所述,如何提供一种通用的语音识别框架,以提升语音识别准确度,成为本领域技术人员急需解决的问题。

发明内容

[0005] 本申请提供语音交互系统,以解决现有技术存在的语音识别准确度较低的问题。本申请另外提供语音交互方法和装置,电视节目播放方法和装置,会议记录方法和装置,语

音识别模型构建方法和装置,智能音箱,智能电视,以及电子设备。

[0006] 本申请提供一种语音交互系统,包括:

[0007] 智能音箱,用于采集目标用户的语音数据,将所述语音数据发送至服务端;

[0008] 服务端,用于构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0009] 本申请还提供一种语音交互方法,包括:

[0010] 构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;

[0011] 针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;

[0012] 根据所述文本序列,执行语音交互处理。

[0013] 可选的,所述通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列,包括:

[0014] 通过所述语音识别模型包括的语言模型,确定与所述语音数据对应的第一文本特征;

[0015] 根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征;

[0016] 至少根据第二文本特征,确定所述文本序列。

[0017] 可选的,所述根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征,包括:

[0018] 通过所述语音识别模型包括的指示器评分模型,根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征。

[0019] 可选的,通过所述指示器评分模型,确定与所述第一文本特征对应的词与各个语言知识的相关度;

[0020] 至少根据与所述相关度大于相关度阈值的词相关的语言知识,确定第二文本特征。

[0021] 可选的,所述方法还包括:

[0022] 通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;

[0023] 将语言知识的编码数据存储至所述语音识别模型包括的语言知识存储模块;

[0024] 所述根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征,包括:

[0025] 根据所述语言知识存储模块存储的所述编码数据和所述第一文本特征,确定第二文本特征。

[0026] 可选的,还包括:

[0027] 从训练数据集中学习得到所述语音识别模型;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息。

[0028] 可选的,所述训练数据中的个性化语言知识库,采用如下方式确定:

[0029] 根据多个训练数据的文本序列标注信息,构建所述个性化语言知识库。

[0030] 可选的,所述个性化语言知识库包括:长尾实体词,反语言模型的实体词,同音不

同字的实体词,上下文中的实体词。

- [0031] 可选的,所述音箱服务领域包括:打电话服务领域;
- [0032] 所述打电话服务领域的语言知识包括:用户通讯录中的人名;
- [0033] 所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:
- [0034] 接收与所述用户对应的智能音箱发送的用户通讯录信息;
- [0035] 将用户通讯录中的人名作为所述用户的个性化语言知识。
- [0036] 可选的,所述音箱服务领域包括:问答服务领域;
- [0037] 所述问答服务领域的语言知识包括:上下文中的文本片段;
- [0038] 所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:
- [0039] 确定上下文文本序列;
- [0040] 将所述上下文文本序列中的文本片段作为所述用户的个性化语言知识。
- [0041] 可选的,所述音箱服务领域包括:多媒体播放服务领域;
- [0042] 所述多媒体播放服务领域的语言知识包括:曲目名;
- [0043] 所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:
- [0044] 确定用户的历史播放节目名;
- [0045] 将所述历史播放节目名作为所述用户的个性化语言知识。
- [0046] 可选的,所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,采用如下方式的至少一个:
- [0047] 根据所述用户的购物数据,确定所述用户的个性化语言知识;
- [0048] 根据所述用户输入的文本信息,确定所述用户的个性化语言知识。
- [0049] 可选的,还包括:
- [0050] 根据交互语音数据,更新所述用户的语言知识库。
- [0051] 本申请还提供一种语音交互方法,包括:
- [0052] 采集目标用户的语音数据;
- [0053] 将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。
- [0054] 本申请还提供一种语音交互方法,包括:
- [0055] 确定用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;
- [0056] 针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列;
- [0057] 根据所述文本序列,执行语音交互处理。
- [0058] 本申请还提供一种电视节目播放方法,包括:
- [0059] 确定用户的个性化节目播放语言知识库;
- [0060] 针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和

所述语言知识库,确定与所述语音指令数据对应的目标节目名;

[0061] 根据所述目标节目名,播放目标节目对象。

[0062] 可选的,还包括:

[0063] 将所述用户的历史播放节目对象的节目名、演员名和/或导演名作为所述用户的个性化节目播放语言知识。

[0064] 可选的,所述根据所述目标节目名,播放目标节目对象,包括:

[0065] 根据节目表,确定与所述目标节目名对应的电视频道和播放时间;

[0066] 根据所述播放时间和所述电视频道,确定目标节目对象;

[0067] 播放所述目标节目对象。

[0068] 可选的,所述根据所述播放时间和所述电视频道,确定目标节目对象,包括:

[0069] 显示与所述目标节目名对应的至少一个电视频道在多个时间播放的多个节目对象;

[0070] 将用户指定的节目对象作为目标节目对象。

[0071] 可选的,还包括:

[0072] 若所述节目表不包括所述目标节目名,则确定与所述目标节目名相关的节目名;

[0073] 显示相关节目名;

[0074] 若用户指定播放相关节目对象,则播放相关节目对象。

[0075] 本申请还提供一种电视节目播放方法,包括:

[0076] 确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名;

[0077] 根据节目表,确定与所述目标节目名对应的目标节目对象;

[0078] 播放所述目标节目对象。

[0079] 可选的,

[0080] 所述根据节目表,确定与所述目标节目名对应的目标节目对象,包括:

[0081] 根据历史节目表,确定与所述目标节目名对应的历史目标节目对象;

[0082] 根据当前节目表,确定与所述目标节目名对应的当前目标节目对象。

[0083] 本申请还提供一种电视节目播放方法,包括:

[0084] 服务端针对智能电视采集的用户的节目播放语音指令数据,确定与语音指令数据对应的目标节目名;

[0085] 根据节目表,确定与所述目标节目名对应的目标节目对象;

[0086] 通过所述智能电视播放所述目标节目对象。

[0087] 本申请还提供一种电视节目播放方法,包括:

[0088] 智能电视采集用户的节目播放语音指令数据;

[0089] 将所述语音指令数据发送至服务端,以便于服务端确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;

[0090] 播放所述目标节目对象。

[0091] 本申请还提供一种会议记录方法,包括:

[0092] 构建会议领域的语言知识库;

[0093] 采集会议语音数据;

[0094] 通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音

数据对应的文本序列,形成会议记录。

[0095] 可选的,还包括:

[0096] 确定与会议语音数据对应的会议领域。

[0097] 本申请还提供一种会议记录方法,包括:

[0098] 构建各个领域的语言知识库;

[0099] 针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;

[0100] 通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0101] 本申请还提供一种会议记录方法,包括:

[0102] 采集目标会议的语音数据;

[0103] 将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0104] 本申请还提供一种语音识别模型构建方法,包括:

[0105] 确定训练数据集;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息;

[0106] 构建端到端的语音识别模型的网络结构;

[0107] 从训练数据集中学习得到所述语音识别模型。

[0108] 可选的,所述模型包括:声音编码器、解码器、语言知识编码器和分类器。

[0109] 可选的,所述模型包括:声音编码器、解码器、语言知识编码器、语言模型、特征融合模块和分类器。

[0110] 可选的,所述模型包括:声音编码器、语言模型、语言知识编码器、指示器评分模型、特征融合模块和分类器。

[0111] 本申请还提供一种语音识别方法,包括:

[0112] 构建用户的包括至少一个领域语言知识的个性化语言知识库;

[0113] 针对终端设备采集的用户语音数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音数据对应的文本序列。

[0114] 本申请还提供一种语音交互装置,包括:

[0115] 知识库构建单元,用于构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;

[0116] 语音识别单元,用于针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;

[0117] 指令处理单元,用于根据所述文本序列,执行语音交互处理。

[0118] 本申请还提供一种电子设备,包括:

[0119] 处理器和存储器;

[0120] 所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所

述文本序列,执行语音交互处理。

[0121] 本申请还提供一种语音交互装置,包括:

[0122] 语音数据采集单元,用于采集目标用户的语音数据;

[0123] 语音数据发送单元,用于将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0124] 本申请还提供一种智能音箱,包括:

[0125] 处理器和存储器;

[0126] 所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标用户的语音数据;将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0127] 本申请还提供一种电子设备,包括:

[0128] 处理器和存储器;

[0129] 所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标用户的语音数据;将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0130] 本申请还提供一种语音识别模型构建装置,包括:

[0131] 数据准备单元,用于确定训练数据集;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息;

[0132] 网络构建单元,用于构建端到端的语音识别模型的网络结构;

[0133] 网络训练单元,用于从训练数据集中学习得到所述语音识别模型。

[0134] 本申请还提供一种电子设备,包括:

[0135] 处理器和存储器;

[0136] 所述存储器,用于存储实现语音识别模型构建方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定训练数据集;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息;构建端到端的语音识别模型的网络结构;从训练数据集中学习得到所述语音识别模型。

[0137] 本申请还提供一种语音交互装置,包括:

[0138] 知识库构建单元,用于确定用户的包括至少一个领域语言知识的个性化语言知识库;

[0139] 语音识别单元,用于针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列;

[0140] 指令处理单元,用于根据所述文本序列,执行语音交互处理。

[0141] 本申请还提供一种智能音箱,包括:

[0142] 处理器和存储器;

[0143] 所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0144] 本申请还提供一种电子设备,包括:

[0145] 处理器和存储器;

[0146] 所述存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定用户的包括至少一个领域语言知识的个性化语言知识库;针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0147] 本申请还提供一种电视节目播放装置,包括:

[0148] 知识库确定单元,用于智能电视确定用户的个性化节目播放语言知识库;

[0149] 节目识别单元,用于针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名;

[0150] 节目播放单元,用于根据所述目标节目名,播放目标节目对象。

[0151] 本申请还提供一种智能电视,包括:

[0152] 处理器和存储器;

[0153] 所述存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定用户的个性化节目播放语言知识库;针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名;根据所述目标节目名,播放目标节目对象。

[0154] 本申请还提供一种电视节目播放装置,包括:

[0155] 节目名识别单元,用于确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名;

[0156] 节目对象确定单元,用于根据节目表,确定与所述目标节目名对应的目标节目对象;

[0157] 播放单元,用于播放所述目标节目对象。

[0158] 本申请还提供一种智能电视,包括:

[0159] 处理器和存储器;

[0160] 所述存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象。

[0161] 本申请还提供一种电视节目播放装置,包括:

[0162] 节目名识别单元,用于针对智能电视采集的用户的节目播放语音指令数据,确定与语音指令数据对应的目标节目名;

- [0163] 节目对象确定单元,用于根据节目表,确定与所述目标节目名对应的目标节目对象;
- [0164] 播放单元,用于通过所述智能电视播放所述目标节目对象。
- [0165] 本申请还提供一种电子设备,包括:
- [0166] 处理器和存储器;
- [0167] 所述存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:针对智能电视采集的用户的节目播放语音指令数据,确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;通过所述智能电视播放所述目标节目对象。
- [0168] 本申请还提供一种电视节目播放装置,包括:
- [0169] 语音指令采集单元,用于智能电视采集用户的节目播放语音指令数据;
- [0170] 语音指令发送单元,用于将所述语音指令数据发送至服务端,以便于服务端确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;
- [0171] 播放单元,用于播放所述目标节目对象。
- [0172] 本申请还提供一种智能电视,包括:
- [0173] 处理器和存储器;
- [0174] 所述存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集用户的节目播放语音指令数据;将所述语音指令数据发送至服务端,以便于服务端确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象。
- [0175] 本申请还提供一种会议记录装置,包括:
- [0176] 知识库构建单元,用于构建会议领域的语言知识库;
- [0177] 语音数据采集单元,用于采集会议语音数据;
- [0178] 语音转写单元,用于通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录。
- [0179] 本申请还提供一种电子设备,包括:
- [0180] 处理器和存储器;
- [0181] 所述存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建会议领域的语言知识库;采集会议语音数据;通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录。
- [0182] 本申请还提供一种会议记录装置,包括:
- [0183] 知识库构建单元,用于构建各个领域的语言知识库;
- [0184] 会议领域确定单元,用于针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;
- [0185] 语音转写单元,用于通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。
- [0186] 本申请还提供一种电子设备,包括:

[0187] 处理器和存储器;

[0188] 所述存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建各个领域的语言知识库;针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0189] 本申请还提供一种会议记录装置,包括:

[0190] 语音数据采集单元,用于采集目标会议的语音数据;

[0191] 语音数据发送单元,用于将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0192] 本申请还提供一种电子设备,包括:

[0193] 处理器和存储器;

[0194] 所述存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标会议的语音数据;将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0195] 本申请还提供一种语音识别装置,包括:

[0196] 知识库构建单元,用于构建用户的包括至少一个领域语言知识的个性化语言知识库;

[0197] 模型预测单元,用于针对终端设备采集的用户语音数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音数据对应的文本序列。

[0198] 本申请还提供一种电子设备,包括:

[0199] 处理器和存储器;

[0200] 所述存储器,用于存储实现语音识别方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建用户的包括至少一个领域语言知识的个性化语言知识库;针对终端设备采集的用户语音数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音数据对应的文本序列。

[0201] 本申请还提供一种计算机可读存储介质,所述计算机可读存储介质中存储有指令,当其在计算机上运行时,使得计算机执行上述各种方法。

[0202] 本申请还提供一种包括指令的计算机程序产品,当其在计算机上运行时,使得计算机执行上述各种方法。

[0203] 与现有技术相比,本申请具有以下优点:

[0204] 本申请实施例提供的语音交互系统,通过智能音箱采集目标用户的语音数据,将所述语音数据发送至服务端;服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。采用这种处理方式,使得至少达到以下技术效果:

[0205] 1) 针对传统技术中,子语言模型在更新语言知识时,需要重新训练,因而花费较多时间,从而影响语音识别实时性的问题,由于本申请提供的系统能够实时更新知识库内容,

并根据实时更新的知识库进行语音识别,因此可以有效兼顾较高的语音识别实时性和准确度。

[0206] 2) 针对传统技术中存在的句式依赖问题,如“打电话给”郑梓豪这个句式,由于本申请提供的系统构建在由词汇构成的语言知识库的基础上,并不依赖句式,某个音箱技能中的词汇可应用在其它音箱技能的语音识别中,因此可以有效提升语音识别的准确度。

[0207] 3) 本申请提供的系统依赖的模型是端到端模型,避免出现传统的非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0208] 4) 针对传统技术中,子语言模型只对相应的音箱技能有效的问题,由于本申请提供的系统构建在由词汇构成的语言知识库的基础上,该知识库可包括多个音箱技能的个性化语言知识,一个音箱技能可参考其它音箱技能的个性化语言知识,因此可以有效提升语音识别的准确度。

[0209] 本申请实施例提供的语音识别模型构建方法,通过确定训练数据集;所述训练数据集包括:语音数据、个性化语言知识库、文本序列标注信息;构建端到端的语音识别模型的网络结构;从训练数据集中学习得到所述语音识别模型;这种处理方式,使得所述模型可引入用户个性化的先验语言知识;因此,可以有效提升模型准确度。同时,该模型与句式无关,因此可以有效提升语音识别的可扩展性。此外,该模型可根据实时更新的语言知识库进行语音识别;因此,可以有效提升语音识别的实时性。同时,该模型可为各种音箱技能执行语音识别处理;因此,可以有效提升语音识别的可扩展性。此外,端到端模型不存在误差传播问题,因此可以有效提升语音识别的准确度。

[0210] 本申请实施例提供的电视节目播放方法,通过智能电视确定用户的个性化节目播放语言知识库;针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名;根据所述目标节目名,播放目标节目对象;这种处理方式,使得能够实时更新节目点播知识库内容,并根据实时更新的知识库进行节目点播语音识别,因此可以有效兼顾较高的语音识别实时性和准确度。同时,该方法依赖的模型是端到端模型,可避免出现非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0211] 本申请实施例提供的电视节目播放方法,通过智能电视确定与节目播放语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象;这种处理方式,使得能够根据用户的节目播放语音指令,自动播放相应节目对象,避免用户逐个频道转换,通过人工方式查找其感兴趣的电视节目;因此,可以有效提升用户体验,以及,节约节目服务端的计算资源。

[0212] 本申请实施例提供的会议记录方法,通过构建会议领域的语言知识库;采集会议语音数据;通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录;这种处理方式,使得能够根据会议所属领域的语言知识库进行会议语音识别,避免传统技术中各个会议领域的子语言模型在更新语言知识时,需要重新训练,因而花费较多时间,从而影响语音识别实时性的问题;因此可以有效兼顾较高的语音识别实时性和准确度。同时,该方法依赖的模型是端到端模型,可避免出现非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

附图说明

- [0213] 图1本申请提供一种语音交互系统的实施例的结构示意图；
- [0214] 图2本申请提供一种语音交互系统的实施例的场景示意图；
- [0215] 图3本申请提供一种语音交互系统的实施例的设备交互示意图；
- [0216] 图4本申请提供一种语音交互系统的实施例的端到端语音识别模型示意图；
- [0217] 图5本申请提供一种语音交互系统的实施例的又一端到端语音识别模型示意图；
- [0218] 图6本申请提供一种语音交互系统的实施例的模型具体示意图；
- [0219] 图7本申请提供一种语音交互系统的实施例的指示器评分模型示意图。

具体实施方式

[0220] 在下面的描述中阐述了很多具体细节以便于充分理解本申请。但是本申请能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本申请内涵的情况下做类似推广，因此本申请不受下面公开的具体实施的限制。

[0221] 在本申请中，提供了语音交互系统、方法和装置，电视节目播放方法和装置，会议记录方法和装置，语音识别模型构建方法和装置，智能音箱，智能电视，以及电子设备。在下面的实施例中逐一对各种方案进行详细说明。

[0222] 第一实施例

[0223] 请参考图1，其为本申请的语音交互系统的实施例的示意图。本实施例提供的语音交互系统包括：服务器1和智能音箱2。

[0224] 服务器1，可以是部署在云端服务器上的服务端，也可以是专用于实现语音交互系统的服务器，可部署在数据中心。

[0225] 智能音箱2，可以是家庭消费者用语音进行上网的一个工具，比如点播歌曲、上网购物，或是了解天气预报，它也可以对智能家居设备进行控制，比如打开窗帘、设置冰箱温度、提前让热水器升温等。

[0226] 请参考图2，其为本申请的语音交互系统的场景示意图。服务端1和智能音箱2间可通过网络连接，如智能音箱2可通过WIFI等方式联网，等等。用户与智能音箱之间通过语音方式进行交互。用户向智能音箱2下达语音指令（如打开兔兔灯等，或打电话给某某等），智能音箱将用户语音数据发送至服务端；服务端通过语音识别技术确定该语音数据的文本序列；根据解码得到的文本序列，执行语音交互处理。

[0227] 请参考图3，其为本申请的语音交互系统的设备示意图。在本实施例中，智能音箱，用于采集目标用户的语音数据，将所述语音数据发送至服务端；服务端，用于构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库；通过端到端的语音识别模型和所述目标用户的所述语言知识库，确定与所述语音数据对应的文本序列；根据所述文本序列，执行语音交互处理。

[0228] 本申请实施例提供的所述系统，可为每个用户构建其专有的语言知识库，如用户A的知识库、用户B的知识库等。所述个性化语言知识库，包括一个用户的至少一个音箱服务领域的词汇，该词汇可与句式无关，不存在句式依赖问题。

[0229] 具体实施时，可以为每个音箱服务构建专用的语言知识库，如为用户A构建“点歌”

技能的语言知识库、“你猜我想”技能的语言知识库、“打电话”技能的语言知识库。采用这种处理方式,使得不同音箱技能可具有不同的语言知识,如用户A的“点歌”技能的语言知识库包括词汇“纪念”、用户A的“打电话”技能的语言知识库包括人名“纪念”,这样就可以提升每种音箱技能的语音识别准确度,不会相互干扰。

[0230] 具体实施时,也可以构建一个多个音箱服务通用的语言知识库,这样在所述语言知识库中就会包括多个音箱服务领域的词汇,如包括“宦妃权倾天下”、“梓豪”、“情人结”等;这种处理方式,使得知识库可包括多个音箱技能的个性化语言知识,一个音箱技能可参考其它音箱技能的个性化语言知识,因此可以有效提升语音识别的准确度,特别是新技能的语音识别准确度。

[0231] 所述语言知识库中的词汇,包括但不限于:长尾实体词(如有声书名《宦妃权倾天下》),反语言模型的实体词(如音乐歌名“情人结”),同音不同字的实体词(如“梓豪”vs“子豪”),还可包括语音交互过程中上下文包括的实体词(如“有角”)。

[0232] 在一个示例中,所述音箱服务领域包括:打电话服务领域;所述打电话服务领域的语言知识包括:用户通讯录中的人名,如“梓豪”;所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:接收与所述用户对应的智能音箱发送的用户通讯录信息;将用户通讯录中的人名作为所述用户的个性化语言知识。采用这种处理方式,使得语言知识库包括通讯录的人名信息,这样就可以更加准确地确定通话对象;因此,可以有效提升通话精准度。

[0233] 在一个示例中,所述音箱服务领域包括:问答服务领域,如天猫精灵的“你想我猜”技能;所述问答服务领域的语言知识包括:上下文中的文本片段;所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:确定与用户的上下文语音数据对应的上下文文本序列,如天猫精灵问到“它有角吗”;将所述上下文文本序列中的文本片段作为所述用户的个性化语言知识,如“有角”这个词。采用这种处理方式,使得语言知识库包括问答过程中的上下文实体信息,这样就可以更加准确地确定答复文本;因此,可以有效提升问答准确度,使得问答过程可以顺利进行。

[0234] 在一个示例中,所述音箱服务领域包括:多媒体播放服务领域,如天猫精灵的“点歌”技能;所述多媒体播放服务领域的语言知识包括:曲目名;所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,包括:确定用户的历史播放节目名;将所述历史播放节目名作为所述用户的个性化语言知识。采用这种处理方式,使得语言知识库包括歌曲名、电影名等,这样就可以更加准确地确定用户想要播放的节目;因此,可以有效提升节目播放准确度,从而提升用户体验。

[0235] 在一个示例中,所述构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库,采用如下方式的至少一个:根据所述用户的购物数据,确定所述用户的个性化语言知识,如用户A通过音箱查询线上订单的物流状态,则将该订单涉及的商品名、品牌名、地址名等作为该用户的个性化语言知识;根据所述用户输入的文本信息,确定所述用户的个性化语言知识,如用户A在带屏音箱上输入“兔兔灯”,则将该词作为该用户的个性化语言知识。采用这种处理方式,可以获得用户的更为丰富的个性化语言知识;因此,可以有效提升语音识别的准确度。

[0236] 在一个示例中,服务端还可用于根据交互语音数据,更新所述用户的语言知识库,

如将用户说到的“金海心”添加到知识库中,或者将天猫精灵说到的“有角”这个词添加到知识库中。采用这种处理方式,使得实时更新语言知识库的知识,因而可以有效提升语音识别的准确度。

[0237] 所述端到端(End2End)的语音识别模型,将语音信号转成文本,该模型采用将声学模型和语言模型合二为一的语音识别框架,这样就不存在模块间的误差传播效应,可以提升语音识别的准确度。

[0238] 在一个示例中,所述端到端的语音识别模型,可以是如图4所示的CLAS(Listen Attend and Spell)模型,包括声音编码器、解码器、语言知识编码器,还可包括分类器。在该语音识别模型中,语言模型并非显式可见。为了引入先验知识(用户的个性化语言知识),以把模型的搜索范围限定到一个合适的范围,就需要引入相关误差,具体实施时可以引入特殊符号(如\$)。在本实施例中,在解码输出上,若该语音识别模型从内存Memory中读取一个词组,会有一个特殊符号「\$」接在对应的词组后边,以此方式来引入解码器decoder的误差,而正常解码没有。

[0239] 本申请实施例提供的所述系统,也可以采用图4以外的其它结构的端到端的语音识别模型,如图5或图6的模型等。在本实施例中,服务端具体用于通过所述语音识别模型包括的语言模型,确定与所述语音数据对应的第一文本特征;根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征;至少根据第二文本特征,确定所述文本序列。

[0240] 所述第一文本特征,包括通过所述语言模型确定的与所述语音数据对应的文本特征。例如,发音“tu tu”的第一文本特征包括“土兔”的文本特征;发音“ji nian”的第一文本特征包括“纪念”的文本特征。

[0241] 所述第二文本特征,包括根据所述语言知识库和所述第一文本特征确定的与所述语音数据对应的文本特征。例如,语言知识库包括“兔兔灯”这个词,发音“tu tu”的第二文本特征包括“兔兔”的文本特征;语言知识库包括“记念”这个词,发音“ji nian”的第二文本特征包括“记念”的文本特征。

[0242] 在一个示例中,所述端到端的语音识别模型可以是如图5所示的模型,包括声音编码器、解码器、语言知识编码器、语言模型、特征融合模块和分类器。在该语音识别模型中,包括显式的语言模型,该语音识别模型可通过特征融合模块对语言模型输出的特征、解码器输出的特征和语言知识编码器输出的特征执行特征融合处理,确定出与用户语音数据对应的文本序列的特征,再通过分类器根据融合的文本特征,确定与用户语音数据对应的文本序列。

[0243] 如图5所示,所述端到端的语音识别模型包括的语言模型,可以是一个通用的语言模型。这样,一方面可根据丰富的语料训练该语言模型,训练样本通常包括常见的语料,可不包括用户的个性化语言知识,如不包括长尾实体“宦妃权倾天下”、反语言模型的“情人结”、同音不同字的“梓豪”。采用这种处理方式,使得用户个性化的语言知识不需要依赖语言模型引入到语音识别中,用户个性化的语言知识不会对语言模型产生影响,当用户出现新的个性化语言知识时,无需根据新知识重新训练语言模型,而是将新增的个性化语言知识增加至该用户的个性化语言知识库中,通过语言知识编码器将用户的实时更新的个性化语言知识引入到语音识别模型中,通过语音识别模型确定文本;因此,可以有效兼顾较高的

语音识别实时性和准确度。另一方面,相对图4所示的模型,由于图5所示的模型可根据丰富的训练语料训练得到准确度更高的语言模型;因此,可以有效提升语音识别准确度。

[0244] 在另一个示例中,所述端到端的语音识别模型可以是如图6所示的模型,包括声音编码器、语言模型、语言知识编码器、指示器评分模型、特征融合模块和分类器。在该语音识别模型中,同样包括显式的语言模型。服务端具体用于根据所述第一文本特征和第二文本特征,确定所述文本序列。如图6所示,服务端可先通过语言知识编码器,对语言知识库中的词汇进行编码,并将编码结果存储再外部内存(External Memory)中;针对待解码的语音数据,可先通过语音识别模型包括的声音编码器对模型的输入数据声学特征进行编码,可将编码结果存储在内存memory中,作为显式的语言模型的输入数据;然后,通过语言模型,根据编码结果确定第一文本特征,如“打开土兔”的特征;然后,再通过指示器评分模型,根据外部内存中的语言知识、和第一文本特征,确定第二文本特征,如“打开兔兔”的特征;然后,再通过语音识别模型包括的特征融合模块融合两路输入:所述第一文本特征和第二文本特征,该模块可将两个特征连接(concatenate)或者将两个特征相加等,如特征融合模块的输出包括“打开兔兔”的特征;然后,根据融合后的文本特征,确定所述文本序列。采用这种处理方式,使得可根据第一文本特征和第二文本特征,确定文本序列;因此,可以有效提升语音识别准确度。

[0245] 在本实施例中,所述服务端要通过指示器评分模型,根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征,可采用如下方式实现:确定与所述第一文本特征对应的词与各个语言知识的相关度;至少根据与所述相关度大于相关度阈值的词相关的语言知识,确定第二文本特征。采用这种处理方式,使得在引入先验知识时,无需加入特殊符号,因而不会改变语言模型的训练数据,轻量化地实现更好的解码;因此,可以有效提升语音识别的准确度。

[0246] 具体实施时,可以是针对第一文本特征的文本中的各个字,根据以当前字结尾的文本片段的第一文本特征(如“打开土兔”的特征),确定以当前字结尾的文本片段中的词与各个语言知识的相关度,如土兔与兔兔灯的相关度;若所述相关度大于相关度阈值,则根据相关的语言知识(如兔兔灯),确定以当前字结尾的文本片段的第二文本特征(如“打开兔兔”的特征)。

[0247] 如图7所示,在解码过程中的每一步,都可对外部内存(External Memory)中的M个词汇候选集(个性化语言知识词汇)进行二分类,如相关或不相关。由图7可见,评分模型包括相关性矩阵为 $B \times T \times M$,T可以是一个矩阵,T矩阵元素包括第一文本特征对应的各个字的编码,M也可以是一个矩阵,M矩阵元素包括知识库中各个词汇的编码。例如,解码第一步处理的文本为“打”,确定该字与知识库中的各个词汇的相关度,没有找到相关度大于相关度阈值的词汇,因此确定相应的第二文本特征包括“打”字的特征,再如解码第四步处理的文本为“打开土兔”,确定其中各个词分别与知识库中的各个词汇的相关度,其中“土兔”与“兔兔灯”间的相关度大于相关度阈值,由此确定相应的第二文本特征包括“打开兔兔”的特征。

[0248] 如图6所示,所述系统可基于全Transformer架构实现端到端语音识别模型,该ASR模型包括未经改造的语言模型,语言模型中没有特殊符号的概念,无需改动训练数据引入特殊符号,因此不会影响语言模型的效果。同时,在不改造语言模型的情况下,通过指示器评分模型,轻量化地引入用户个性化语言知识的先验知识,把模型的搜索范围限定到一个

合适的范围;并且是在全神经网络的架构上,端到端地实现这个特性。因此,可以轻量化地实现更好的解码。

[0249] 具体实施时,为了引入先验知识,需要引入相关误差,可以是在图5所示的模型基础上使用特殊符号(如\$)。采用这种处理方式,使得在引入先验知识时会改变语言模型的训练数据,从而使得识别出的句子中会随机出现\$。

[0250] 具体实施时,可以是图5或图6所示的模型基础上不使用特殊符号,而是通过指示器评分模型引入先验知识。通过指示器评分模型,可结合外部内存中的语言知识、和第一文本特征,确定第二文本特征,如第一文本特征为“打开土兔”的特征,知识库中包括“兔兔灯”,第二文本特征为“打开兔兔”的特征。采用这种处理方式,使得在主模型(包括语言模型等)能训练的很好的情况下,保证语音识别模型一定会从外部内存中读取信息,也就是说,在损失值Loss的设计上,把语言知识编码器biasEncoder的Loss引入到整个解码Decoder的损失Loss之中。

[0251] 在本实施例中,服务端还用于通过所述语音识别模型包括的语言知识编码器,对语言知识执行编码处理;将语言知识的编码数据存储至所述语音识别模型包括的语言知识存储模块(即上述外部内存);并具体用于通过指示器评分模型,根据所述语言知识存储模块存储的所述编码数据和所述第一文本特征,确定第二文本特征。

[0252] 所述服务端确定与所述语音数据对应的文本序列之后,可根据所述文本序列,执行语音交互处理。例如,在问答场景(如你猜我想)中,可根据用户提供的信息,确定回复信息,将回复信息发送至智能音箱,展示给用户,如回复“有角”;或者,在点播场景中,可将用户点播的歌曲数据发送至智能音箱播放,如播放“记念”;或者,在打电话场景中,可为智能音箱与目标通过人的通讯设备建立连接,如拨通“梓豪”的手机。

[0253] 至此,对端到端语音识别模型预测阶段的处理过程进行了说明。下面对该模型的训练阶段的处理方式进行说明。

[0254] 在端到端语音识别模型训练阶段,服务端用于确定训练数据集;所述训练数据可包括:语音数据、个性化语言知识库、文本序列标注信息,还可包括个性化语言知识是否有效的信息等;从所述训练数据集中学习得到所述模型。在训练所述模型时,可将语音数据的声学特征作为模型的输入数据,将文本序列标注信息作为输出数据,基于个性化语言知识库构建指示器评分模型的M矩阵,采用机器学习方式从训练数据集中学习得到所述模型。

[0255] 在一个示例中,所述训练数据中的个性化语言知识库,采用如下方式确定:根据多个训练数据的文本序列标注信息,自动构建所述个性化语言知识库。采用这种处理方式,使得不需要额外的数据,就可以生成大量的训练数据;因此,可以有效提升端到端语音识别模型的构建效率。

[0256] 具体实施时,服务端还用于构建语言知识编码器。语言知识编码器的输入数据可以为一组词组的文本,输出数据为对应词组的向量化编码;这些编码中包含对这些词的信息;这些知识信息可帮助解码器模块准确地输出文本,如“情人结”、“打开兔兔”、“梓豪”等。在本实施例中,训练数据包括语音及其对应的文本,因此可以从文本中通过随机截取文本片段,作为词组,进而作为语言知识编码器的输入;这些词组,结合对应的语音信号,可以帮助解码器更准确地解码出文本。在训练过程中,解码器就学会了如何从知识库中选择相关的知识来辅助解码;整个训练/学习过程,均是端到端的,不需要额外训练数据。

[0257] 从上述实施例可见,本申请实施例提供的语音交互系统,通过智能音箱采集目标用户的语音数据,将所述语音数据发送至服务端;服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。采用这种处理方式,使得至少达到以下技术效果:

[0258] 1) 针对传统技术中,子语言模型在更新语言知识时,需要重新训练,因而花费较多时间,从而影响语音识别实时性的问题,由于本申请提供的系统能够实时更新知识库内容,并根据实时更新的知识库进行语音识别,因此可以有效兼顾较高的语音识别实时性和准确度。

[0259] 2) 针对传统技术中存在的句式依赖问题,如“打电话给”郑梓豪这个句式,由于本申请提供的系统构建在由词汇构成的语言知识库的基础上,并不依赖句式,某个音箱技能中的词汇可应用在其它音箱技能的语音识别中,因此可以有效提升语音识别的准确度。

[0260] 3) 本申请提供的系统依赖的模型是端到端模型,避免出现传统的非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0261] 4) 针对传统技术中,子语言模型只对相应的音箱技能有效的问题,由于本申请提供的系统构建在由词汇构成的语言知识库的基础上,该知识库可包括多个音箱技能的个性化语言知识,一个音箱技能可参考其它音箱技能的个性化语言知识,因此可以有效提升语音识别的准确度。

[0262] 第二实施例

[0263] 在上述的实施例中,提供了一种语音交互系统,与之相对应的,本申请还提供一种语音交互方法,该方法的执行主体可以是服务器,也可以是能够实现所述方法的任何设备。该方法是与上述系统的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0264] 在本实施例中,所述方法可包括如下步骤:

[0265] 步骤1:构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库。

[0266] 所述个性化语言知识库,包括但不限于:长尾实体词,反语言模型的实体词,同音不同字的实体词,上下文中的实体词。

[0267] 在一个示例中,所述音箱服务领域包括:打电话服务领域;所述打电话服务领域的语言知识包括:用户通讯录中的人名;在这种情况下,步骤1可包括如下子步骤:1.1) 接收与所述用户对应的智能音箱发送的用户通讯录信息;1.2) 将用户通讯录中的人名作为所述用户的个性化语言知识。

[0268] 在一个示例中,所述音箱服务领域包括:问答服务领域;所述问答服务领域的语言知识包括:上下文中的文本片段;在这种情况下,步骤1可包括如下子步骤:1.3) 确定上下文文本序列;1.4) 将所述上下文文本序列中的文本片段作为所述用户的个性化语言知识。

[0269] 在一个示例中,所述音箱服务领域包括:多媒体播放服务领域;所述多媒体播放服务领域的语言知识包括:曲目名;在这种情况下,步骤1可包括如下子步骤:1.5) 确定用户的历史播放节目名;1.6) 将所述历史播放节目名作为所述用户的个性化语言知识。

[0270] 在一个示例中,步骤1可采用如下方式的至少一个:1.7) 根据所述用户的购物数

据,确定所述用户的个性化语言知识;1.8)根据所述用户输入的文本信息,确定所述用户的个性化语言知识。

[0271] 在一个示例中,所述方法还可包括如下步骤:根据交互语音数据,更新所述用户的语言知识库。

[0272] 步骤2:针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列。

[0273] 在一个示例中,所述语音识别模型包括语言模型,步骤2可包括如下子步骤:2.1)通过所述语言模型,确定与所述语音数据对应的第一文本特征;2.2)根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征;2.3)至少根据第二文本特征,确定所述文本序列。

[0274] 在一个示例中,所述语音识别模型包括指示器评分模型,步骤2.2可采用如下方式实现:通过所述指示器评分模型,根据所述语言知识库和所述第一文本特征,确定与所述语音数据对应的第二文本特征。

[0275] 在一个示例中,步骤2.2可包括如下子步骤:2.2.1)通过所述指示器评分模型,确定与所述第一文本特征对应的词与各个语言知识的相关度;至少根据与所述相关度大于相关度阈值的词相关的语言知识,确定第二文本特征。

[0276] 在一个示例中,所述语音识别模型包括语言知识编码器;相应的,所述方法还可包括如下步骤:通过所述语言知识编码器,对语言知识执行编码处理;将语言知识的编码数据存储至所述语音识别模型包括的语言知识存储模块;相应的,步骤2.2可采用如下方式实现:根据所述语言知识存储模块存储的所述编码数据和所述第一文本特征,确定第二文本特征。

[0277] 步骤3:根据所述文本序列,执行语音交互处理。

[0278] 在一个示例中,所述方法还可包括如下步骤:确定训练数据集;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息;从所述训练数据集中学习得到所述语音识别模型。

[0279] 在一个示例中,所述训练数据中的个性化语言知识库,可采用如下方式确定:根据多个训练数据的文本序列标注信息,构建所述个性化语言知识库。

[0280] 从上述实施例可见,本申请实施例提供的语音交互方法,通过构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。采用这种处理方式,使得至少达到以下技术效果:

[0281] 1)针对传统技术中,子语言模型在更新语言知识时,需要重新训练,因而花费较多时间,从而影响语音识别实时性的问题,由于本申请提供的方法能够实时更新知识库内容,并根据实时更新的知识库进行语音识别,因此可以有效兼顾较高的语音识别实时性和准确度。

[0282] 2)针对传统技术中存在的句式依赖问题,如“打电话给”郑梓豪这个句式,由于本申请提供的方法构建在由词汇构成的语言知识库的基础上,并不依赖句式,某个音箱技能中的词汇可应用在其它音箱技能的语音识别中,因此可以有效提升语音识别的准确度。

[0283] 3) 本申请提供的方法依赖的模型是端到端模型,避免出现传统的非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0284] 4) 针对传统技术中,子语言模型只对相应的音箱技能有效的问题,由于本申请提供的方法构建在由词汇构成的语言知识库的基础上,该知识库可包括多个音箱技能的个性化语言知识,一个音箱技能可参考其它音箱技能的个性化语言知识,因此可以有效提升语音识别的准确度。

[0285] 第三实施例

[0286] 在上述的实施例中,提供了一种语音交互方法,与之相对应的,本申请还提供一种语音交互装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0287] 本申请提供的一种语音交互装置包括:

[0288] 知识库构建单元,用于构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;

[0289] 语音识别单元,用于针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;

[0290] 指令处理单元,用于根据所述文本序列,执行语音交互处理。

[0291] 第四实施例

[0292] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0293] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;针对智能音箱发送的目标用户的语音数据,通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0294] 第五实施例

[0295] 在上述的实施例中,提供了一种语音交互系统,与之相对应的,本申请还提供一种语音交互方法,该方法的执行主体可以是智能音箱,智能电视,聊天机器人,等等。该方法是与上述系统的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0296] 本申请提供的一种语音识别方法,可包括如下步骤:

[0297] 步骤1:采集目标用户的语音数据;

[0298] 步骤2:将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0299] 第六实施例

[0300] 在上述的实施例中,提供了一种语音交互方法,与之相对应的,本申请还提供一种语音交互装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的

部分不再赘述,请参见实施例一中的相应部分。

[0301] 本申请提供一种语音交互装置包括:

[0302] 语音数据采集单元,用于采集目标用户的语音数据;

[0303] 语音数据发送单元,用于将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0304] 第七实施例

[0305] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0306] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现语音交互方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标用户的语音数据;将所述语音数据发送至服务端,以便于服务端生成各个用户的包括至少一个领域语言知识的个性化语言知识库;通过端到端的语音识别模型和所述目标用户的所述语言知识库,确定与所述语音数据对应的文本序列;根据所述文本序列,执行语音交互处理。

[0307] 所述电子设备,可以是智能音箱,智能电视,聊天机器人,等等。

[0308] 在一个示例中,所述电子设备为智能音箱,服务端生成各个用户的包括至少一个音箱服务领域语言知识的个性化语言知识库。

[0309] 第八实施例

[0310] 在上述的实施例中,提供了一种语音交互系统,与之相对应的,本申请还提供一种语音识别模型构建方法,该方法的执行主体可以是服务端等等。该方法是与上述系统的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0311] 本申请提供一种语音识别模型构建方法,可包括如下步骤:

[0312] 步骤1:确定训练数据集。

[0313] 所述训练数据,包括但不限于:语音数据、个性化语言知识库、文本序列标注信息。

[0314] 步骤2:构建端到端的语音识别模型的网络结构。

[0315] 如图4所示,在一个示例中,所述模型可采用CLS网络结构,包括:声音编码器、解码器、语言知识编码器,还可包括分类器。

[0316] 如图5所示,在一个示例中,所述模型可包括:声音编码器、解码器、语言知识编码器、语言模型、特征融合模块和分类器。

[0317] 如图6所示,在一个示例中,所述模型可包括:声音编码器、语言模型、语言知识编码器、指示器评分模型、特征融合模块和分类器。

[0318] 步骤3:从训练数据集中学习得到所述语音识别模型。

[0319] 从上述实施例可见,本申请实施例提供的语音识别模型构建方法,通过确定训练数据集;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息;构建端到端的语音识别模型的网络结构;从训练数据集中学习得到所述语音识别模型;这种处理方式,

使得所述模型可引入用户个性化的先验语言知识;因此,可以有效提升模型准确度。同时,该模型与句式无关,因此可以有效提升语音识别的可扩展性。此外,该模型可根据实时更新的语言知识库进行语音识别;因此,可以有效提升语音识别的实时性。同时,该模型可为各种音箱技能执行语音识别处理;因此,可以有效提升语音识别的可扩展性。此外,端到端模型不存在误差传播问题,因此可以有效提升语音识别的准确度。

[0320] 第九实施例

[0321] 在上述的实施例中,提供了一种语音识别模型构建方法,与之相对应的,本申请还提供一种语音识别模型构建装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0322] 本申请提供一种语音识别模型构建装置,包括:

[0323] 数据准备单元,用于确定训练数据集;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息;

[0324] 网络构建单元,用于构建端到端的语音识别模型的网络结构;

[0325] 网络训练单元,用于从训练数据集中学习得到所述语音识别模型。

[0326] 第十实施例

[0327] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0328] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现语音识别模型构建方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定训练数据集;所述训练数据包括:语音数据、个性化语言知识库、文本序列标注信息;构建端到端的语音识别模型的网络结构;从训练数据集中学习得到所述语音识别模型。

[0329] 第十一实施例

[0330] 在上述的实施例中,提供了一种语音交互系统,与之相对应的,本申请还提供一种语音交互方法,该方法的执行主体可以是智能音箱等等。该方法是与上述系统的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0331] 本申请提供一种语音交互方法包括如下步骤:

[0332] 步骤1:确定用户的包括至少一个领域语言知识的个性化语言知识库;

[0333] 步骤2:针对采集到的所述用户的语音数据,通过端到端的语音识别模型、和所述语言知识库,确定与所述语音数据对应的文本序列;

[0334] 步骤3:根据所述文本序列,执行语音交互处理。

[0335] 本实施例提供的所述方法与实施例一提供的所述系统间的不同之处包括,本实施例提供的所述方法可用于智能音箱端,语音识别处理在端上完成,可以不需要服务端的配合。

[0336] 第十二实施例

[0337] 在上述的实施例中,提供了一种语音交互方法,与之相对应的,本申请还提供一种语音交互装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0338] 本申请提供一种语音交互装置包括：

[0339] 知识库构建单元，用于确定用户的包括至少一个领域语言知识的个性化语言知识库；

[0340] 语音识别单元，用于针对采集到的所述用户的语音数据，通过端到端的语音识别模型、和所述语言知识库，确定与所述语音数据对应的文本序列；

[0341] 指令处理单元，用于根据所述文本序列，执行语音交互处理。

[0342] 第十三实施例

[0343] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0344] 本实施例的一种电子设备，该电子设备包括：处理器和存储器；存储器，用于存储实现语音交互方法的程序，该设备通电并通过所述处理器运行该方法的程序后，执行下述步骤：确定用户的包括至少一个领域语言知识的个性化语言知识库；针对采集到的所述用户的语音数据，通过端到端的语音识别模型、和所述语言知识库，确定与所述语音数据对应的文本序列；根据所述文本序列，执行语音交互处理。

[0345] 第十四实施例

[0346] 在上述的实施例中，提供了一种语音交互系统，与之相对应的，本申请还提供一种电视节目播放方法，该方法的执行主体包括但不限于：智能电视、电视遥控器等。该方法是与上述系统的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述，请参见实施例一中的相应部分。

[0347] 本申请提供一种电视节目播放方法包括如下步骤：

[0348] 步骤1：确定用户的个性化节目播放语言知识库。

[0349] 所述语言知识库，包括但不限于：电视节目名、频道名、演员名等与电视节目播放有关的词汇。

[0350] 在一个示例中，所述方法还可包括如下步骤：将所述用户的历史播放节目对象的节目名、演员名和/或导演名作为所述用户的个性化节目播放语言知识。采用这种处理方式，使得所述语言知识库内的词汇可包括与用户历史观看过的电视节目有关的词汇。

[0351] 步骤2：针对采集到的所述用户的节目播放语音指令数据，通过端到端的语音识别模型和所述语言知识库，确定与所述语音指令数据对应的目标节目名。

[0352] 智能电视可通过遥控器等装置采集用户的节目播放语音指令数据，如用户发出语音指令“我想看哈利波特”，通过步骤2可识别出节目名为“哈利波特”。本步骤的语音识别的具体实施方式可参见实施例一的相关说明，此处不再赘述。

[0353] 在确定目标节目名后，就可以进入下一步播放相关的电视节目。

[0354] 步骤3：根据所述目标节目名，播放目标节目对象。

[0355] 在一个示例中，步骤3可包括如下子步骤：3.1) 根据各个电视频道的节目表（如最近一周的节目表，可包括最近一周播放的节目信息和当前正在播放的节目信息），确定与所述目标节目名相关的电视频道和播放时间；3.2) 根据所述播放时间和所述电视频道，确定目标节目对象；3.3) 播放所述目标节目对象。

[0356] 以某地区有线电视为例，用户对遥控器说“我想看哈利波特”，遥控器首先识别出

用户想要点播或回看的节目名为“哈利波特”,然后,可根据一周内可回看的电视节目表,从中查找哪个频道、何时播放了哈利波特,如果找到了,就可播放这个可回看的节目对象或者相关频道当前正在播放的节目对象。

[0357] 下表示出了本实施例中的节目表。

[0358]	播放时间	电视频道	节目名
	2020/6/1 19:00-19:30	中央1台	新闻联播
	...		
	2020/6/1 20:00-22:30	中央6台	哈利波特1
	...		
	2020/6/3 20:00-22:30	中央6台	哈利波特2
	...		

[0359] 如上表所示,与步骤2确定的目标节目名对应的节目对象可包括多个节目对象,如6月1日和3日均播放了哈利波特。在这种情况下,步骤3.2还可包括如下子步骤:3.2.1)显示与所述目标节目名对应的至少一个电视频道在多个时间播放的多个节目对象;3.2.2)将用户指定的节目对象作为目标节目对象。采用这种处理方式,使得可在电视屏幕上显示所有相关的节目对象,供用户选择,然后播放用户指定的目标节目对象。例如,如果找到了多个可回看的节目,如一个电视剧的很多集,则显示这些可回看的集,让用户选择想要观看哪一集。

[0360] 具体实施时,步骤3还可包括如下子步骤:若所述节目表不包括所述目标节目名,则确定与所述目标节目名相关的节目名;显示相关节目名;若用户指定播放相关节目对象,则播放相关节目对象。采用这种处理方式,使得如果没找到用户先要观看的节目,则还可向用户推荐相关的电视节目,如用户想要看“印象西湖”纪录片,但近一周并没有播放该纪录片,则可以播放有关西湖的其他节目,如“解密灵隐寺”、“西湖十景”、“顾景舟”等节目。

[0361] 从上述实施例可见,本申请实施例提供的电视节目播放方法,通过智能电视确定用户的个性化节目播放语言知识库;针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名;根据所述目标节目名,播放目标节目对象;这种处理方式,使得能够实时更新节目点播知识库内容,并根据实时更新的知识库进行节目点播语音识别,因此可以有效兼顾较高的语音识别实时性和准确度。同时,该方法依赖的模型是端到端模型,可避免出现非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0362] 第十五实施例

[0363] 在上述的实施例中,提供了一种电视节目播放方法,与之相对应的,本申请还提供一种电视节目播放装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0364] 本申请提供的一种电视节目播放装置包括:

[0365] 知识库确定单元,用于智能电视确定用户的个性化节目播放语言知识库;

[0366] 节目识别单元,用于针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名;

[0367] 节目播放单元,用于根据所述目标节目名,播放目标节目对象。

[0368] 第十六实施例

[0369] 本申请还提供一种智能电视。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0370] 本实施例的一种智能电视,该智能电视包括:处理器和存储器;存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:智能电视确定用户的个性化节目播放语言知识库;针对采集到的所述用户的节目播放语音指令数据,通过端到端的语音识别模型和所述语言知识库,确定与所述语音指令数据对应的目标节目名;根据所述目标节目名,播放目标节目对象。

[0371] 第十七实施例

[0372] 在上述的实施例中,提供了一种电视节目播放方法,与之相对应的,本申请还提供一种电视节目播放方法,该方法的执行主体可以是智能电视等。该方法是与上述系统的实施例相对应。本实施例与第十四实施例内容相同的部分不再赘述,请参见实施例十四中的相应部分。

[0373] 本申请提供的一种电视节目播放方法包括如下步骤:

[0374] 步骤1:确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名;

[0375] 步骤2:根据节目表,确定与所述目标节目名对应的目标节目对象;

[0376] 步骤3:播放所述目标节目对象。

[0377] 在一个示例中,步骤2可包括如下子步骤:2.1)根据各个电视频道的历史节目表,确定与所述目标节目名对应的历史目标节目对象;2.2)根据各个电视频道的当前节目表,确定与所述目标节目名对应的当前目标节目对象。

[0378] 从上述实施例可见,本申请实施例提供的电视节目播放方法,通过智能电视确定与节目播放语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象;这种处理方式,使得能够根据用户的节目播放语音指令,自动播放相应节目对象,避免用户逐个频道转换,通过人工方式查找其感兴趣的电视节目;因此,可以有效提升用户体验,以及,节约节目服务端的计算资源。

[0379] 第十八实施例

[0380] 在上述的实施例中,提供了一种电视节目播放方法,与之相对应的,本申请还提供一种电视节目播放装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0381] 本申请提供的一种电视节目播放装置包括:

[0382] 节目名识别单元,用于确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名;

[0383] 节目对象确定单元,用于根据节目表,确定与所述目标节目名对应的目标节目对象;

[0384] 播放单元,用于播放所述目标节目对象。

[0385] 第十九实施例

[0386] 本申请还提供一种智能电视。由于设备实施例基本相似于方法实施例,所以描述

得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0387] 本实施例的一种智能电视,该智能电视包括:处理器和存储器;存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:确定与智能电视采集到的用户的节目播放语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象。

[0388] 第二十实施例

[0389] 在上述的实施例中,提供了一种电视节目播放方法,与之相对应的,本申请还提供一种电视节目播放方法,该方法的执行主体可以是服务端等。该方法是与上述系统的实施例相对应。本实施例与第十七实施例内容相同的部分不再赘述,请参见实施例十七中的相应部分。

[0390] 本申请提供的一种电视节目播放方法包括如下步骤:

[0391] 步骤1:服务端针对智能电视采集的用户的节目播放语音指令数据,确定与语音指令数据对应的目标节目名;

[0392] 步骤2:根据节目表,确定与所述目标节目名对应的目标节目对象;

[0393] 步骤3:通过所述智能电视播放所述目标节目对象。

[0394] 具体实施时,可以是将目标节目对象的视频流发送至智能电视端播放。

[0395] 从上述实施例可见,本申请实施例提供的电视节目播放方法,通过服务端针对智能电视采集的用户的节目播放语音指令数据,确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;通过所述智能电视播放所述目标节目对象;这种处理方式,使得能够根据用户的节目播放语音指令,自动播放相应节目对象,避免用户逐个频道转换,通过人工方式查找其感兴趣的电视节目;因此,可以有效提升用户体验,以及,节约节目服务端的计算资源。

[0396] 第二十一实施例

[0397] 在上述的实施例中,提供了一种电视节目播放方法,与之相对应的,本申请还提供一种电视节目播放装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0398] 本申请提供的一种电视节目播放装置包括:

[0399] 节目名识别单元,用于服务端针对智能电视采集的用户的节目播放语音指令数据,确定与语音指令数据对应的目标节目名;

[0400] 节目对象确定单元,用于根据节目表,确定与所述目标节目名对应的目标节目对象;

[0401] 播放单元,用于通过所述智能电视播放所述目标节目对象。

[0402] 第二十二实施例

[0403] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0404] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行

下述步骤:针对智能电视采集的用户的节目播放语音指令数据,确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;通过所述智能电视播放所述目标节目对象。

[0405] 第二十三实施例

[0406] 在上述的实施例中,提供了一种电视节目播放方法,与之相对应的,本申请还提供一种电视节目播放方法,该方法的执行主体可以是智能电视等。该方法是与上述系统的实施例相对应。本实施例与第十七实施例内容相同的部分不再赘述,请参见实施例十七中的相应部分。

[0407] 本申请提供的一种电视节目播放方法包括如下步骤:

[0408] 步骤1:智能电视采集用户的节目播放语音指令数据;

[0409] 步骤2:将所述语音指令数据发送至服务端,以便于服务端确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;

[0410] 步骤3:播放所述目标节目对象。

[0411] 具体实施时,智能电视可接收服务端发送的目标节目对象的视频流。

[0412] 从上述实施例可见,本申请实施例提供的电视节目播放方法,通过智能电视采集用户的节目播放语音指令数据;将所述语音指令数据发送至服务端,以便于服务端确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象;这种处理方式,使得能够根据用户的节目播放语音指令,自动播放相应节目对象,避免用户逐个频道转换,通过人工方式查找其感兴趣的电视节目;因此,可以有效提升用户体验,以及,节约节目服务端的计算资源。

[0413] 第二十四实施例

[0414] 在上述的实施例中,提供了一种电视节目播放方法,与之相对应的,本申请还提供一种电视节目播放装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0415] 本申请提供的一种电视节目播放装置包括:

[0416] 语音指令采集单元,用于智能电视采集用户的节目播放语音指令数据;

[0417] 语音指令发送单元,用于将所述语音指令数据发送至服务端,以便于服务端确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;

[0418] 播放单元,用于播放所述目标节目对象。

[0419] 第二十五实施例

[0420] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0421] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现电视节目播放方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:智能电视采集用户的节目播放语音指令数据;将所述语音指令数据发送至服务端,以便于服务端确定与语音指令数据对应的目标节目名;根据节目表,确定与所述目标节目名对应的目标节目对象;播放所述目标节目对象。

[0422] 第二十六实施例

[0423] 在上述的实施例中,提供了一种语音交互方法,与之相对应的,本申请还提供一种会议记录方法,该方法的执行主体可以是部署在会议现场的电子设备,如庭审一体机等。该方法是与上述系统的实施例相对应。本实施例与第二实施例内容相同的部分不再赘述,请参见实施例二中的相应部分。

[0424] 本申请提供的一种会议记录方法包括如下步骤:

[0425] 步骤1:构建会议领域的语言知识库。

[0426] 所述会议领域,可以是各种应用领域,如计算机领域、医学领域、法律领域、专利领域等。具体实施时,可以构建多个会议领域的语言知识库,如计算机领域的语言知识库、医学领域的语言知识库、法律领域的语言知识库、专利领域的语言知识库等。

[0427] 具体实施时,针对一个会议领域,可以根据该领域的各种文字资料、多媒体资料,确定该领域的语言知识,形成相应的语言知识库。

[0428] 步骤2:采集会议语音数据。

[0429] 以庭审会议为例,可通过连接庭审一体机的传声器,采集庭审过程中各方人员的语音数据。

[0430] 在一个示例中,所述方法还可包括如下步骤:确定与会议语音数据对应的会议领域。具体实施时,可以由用户指定会议领域,如在启动会议记录时指定会议领域;也可以是通过其它方式自动确定会议领域。

[0431] 步骤3:通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录。

[0432] 所述端到端的语音识别模型,可以从非会议领域的语料库中学习得到的模型,该模型可不包括会议领域的语言知识。为了准确识别会议语音中的专业语言(专业术语等),所述模型需结合所述会议领域的语言知识库,对相应领域的会议语音数据进行语音识别处理,形成会议记录。

[0433] 所述端到端的语音识别模型,可以是如图4、图5和图6所示的模型,该模型包括语言知识编码器。具体实施时,可以由服务端从语料库中学习得到的端到端的语音识别模型,服务端将该模型发送至各个终端设备(如庭审一体机),终端设备包括相关领域的语言知识库,终端设备通过该语音识别模型,并结合相关领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录。

[0434] 从上述实施例可见,本申请实施例提供的会议记录方法,通过构建会议领域的语言知识库;采集会议语音数据;通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录;这种处理方式,使得能够根据会议所属领域的语言知识库进行会议语音识别,避免传统技术中各个会议领域的子语言模型在更新语言知识时,需要重新训练,因而花费较多时间,从而影响语音识别实时性的问题;因此可以有效兼顾较高的语音识别实时性和准确度。同时,该方法依赖的模型是端到端模型,可避免出现非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0435] 第二十七实施例

[0436] 在上述的实施例中,提供了一种会议记录方法,与之相对应的,本申请还提供一种会议记录装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的

部分不再赘述,请参见实施例一中的相应部分。

[0437] 本申请提供一种会议记录装置包括:

[0438] 知识库构建单元,用于构建会议领域的语言知识库;

[0439] 语音数据采集单元,用于采集会议语音数据;

[0440] 语音转写单元,用于通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录。

[0441] 第二十八实施例

[0442] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0443] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建会议领域的语言知识库;采集会议语音数据;通过端到端的语音识别模型和所述会议领域的语言知识库,确定与所述会议语音数据对应的文本序列,形成会议记录。

[0444] 第二十九实施例

[0445] 在上述的实施例中,提供了一种语音交互方法,与之相对应的,本申请还提供一种会议记录方法,该方法的执行主体可以是服务端等。该方法是与上述系统的实施例相对应。本实施例与第二实施例内容相同的部分不再赘述,请参见实施例二中的相应部分。

[0446] 本申请提供一种会议记录方法包括如下步骤:

[0447] 步骤1:构建各个领域的语言知识库。

[0448] 步骤2:针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域。

[0449] 步骤3:通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0450] 从上述实施例可见,本申请实施例提供的会议记录方法,通过构建各个领域的语言知识库;针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录;这种处理方式,使得能够根据会议所属领域的语言知识库进行会议语音识别,避免传统技术中各个会议领域的子语言模型在更新语言知识时,需要重新训练,因而花费较多时间,从而影响语音识别实时性的问题;因此可以有效兼顾较高的语音识别实时性和准确度。同时,该方法依赖的模型是端到端模型,可避免出现非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0451] 第三十实施例

[0452] 在上述的实施例中,提供了一种会议记录方法,与之相对应的,本申请还提供一种会议记录装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0453] 本申请提供一种会议记录装置包括:

[0454] 知识库构建单元,用于构建各个领域的语言知识库;

[0455] 会议领域确定单元,用于针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;

[0456] 语音转写单元,用于通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0457] 第三十一实施例

[0458] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0459] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:构建各个领域的语言知识库;针对终端设备发送的目标会议的语音数据,确定目标会议所属的领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0460] 第三十二实施例

[0461] 在上述的实施例中,提供了一种语音交互方法,与之相对应的,本申请还提供一种会议记录方法,该方法的执行主体可以是终端设备等。该方法是与上述系统的实施例相对应。本实施例与第二实施例内容相同的部分不再赘述,请参见实施例二中的相应部分。

[0462] 本申请提供的一种会议记录方法包括如下步骤:

[0463] 步骤1:采集目标会议的语音数据。

[0464] 步骤2:将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0465] 从上述实施例可见,本申请实施例提供的会议记录方法,通过采集目标会议的语音数据;将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录;这种处理方式,使得能够根据会议所属领域的语言知识库进行会议语音识别,避免传统技术中各个会议领域的子语言模型在更新语言知识时,需要重新训练,因而花费较多时间,从而影响语音识别实时性的问题;因此可以有效兼顾较高的语音识别实时性和准确度。同时,该方法依赖的模型是端到端模型,可避免出现非端到端模型存在的误差传播问题,因此可以有效提升语音识别的准确度。

[0466] 第三十三实施例

[0467] 在上述的实施例中,提供了一种会议记录方法,与之相对应的,本申请还提供一种会议记录装置。该装置是与上述方法的实施例相对应。本实施例与第一实施例内容相同的部分不再赘述,请参见实施例一中的相应部分。

[0468] 本申请提供的一种会议记录装置包括:

[0469] 语音数据采集单元,用于采集目标会议的语音数据;

[0470] 语音数据发送单元,用于将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0471] 第三十四实施例

[0472] 本申请还提供一种电子设备。由于设备实施例基本相似于方法实施例,所以描述

得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0473] 本实施例的一种电子设备,该电子设备包括:处理器和存储器;存储器,用于存储实现会议记录方法的程序,该设备通电并通过所述处理器运行该方法的程序后,执行下述步骤:采集目标会议的语音数据;将所述语音数据发送至服务端,以便于服务端确定目标会议所属领域;通过端到端的语音识别模型和目标会议领域的语言知识库,确定与所述语音数据对应的文本序列,形成目标会议的会议记录。

[0474] 本申请虽然以较佳实施例公开如上,但其并不是用来限定本申请,任何本领域技术人员在不脱离本申请的精神和范围内,都可以做出可能的变动和修改,因此本申请的保护范围应当以本申请权利要求所界定的范围为准。

[0475] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0476] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0477] 1、计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括非暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0478] 2、本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

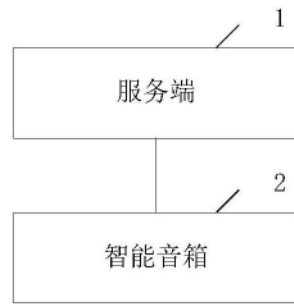


图1

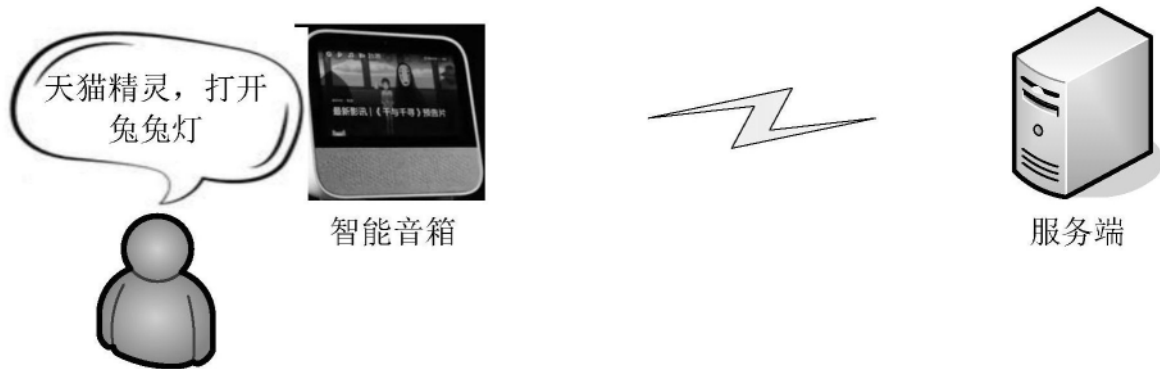


图2

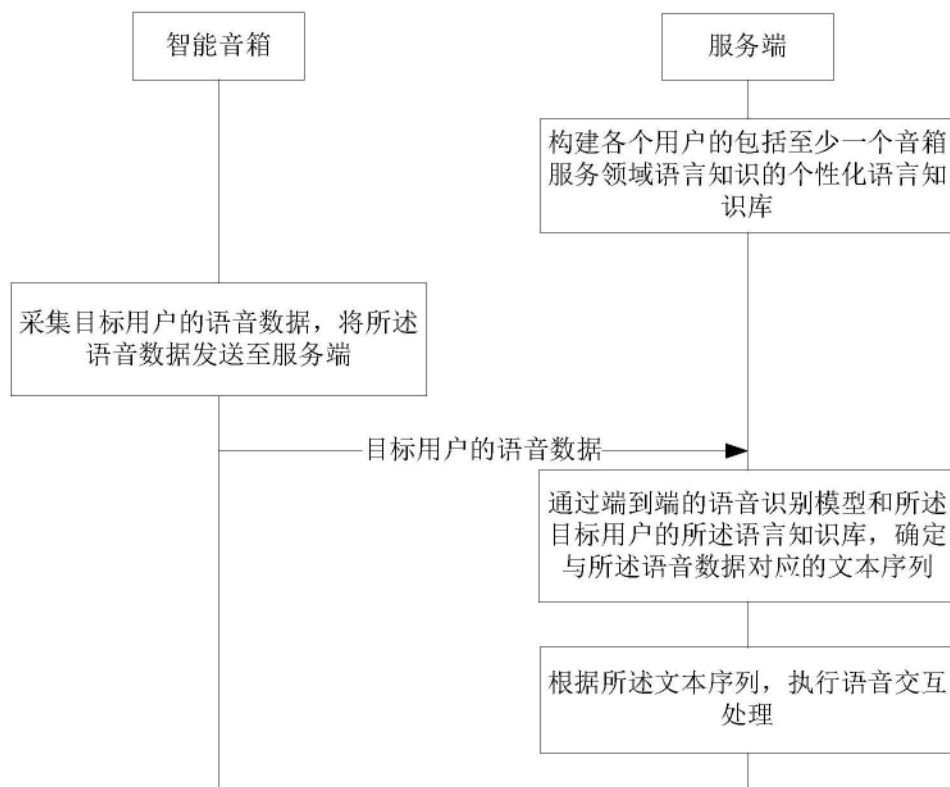


图3

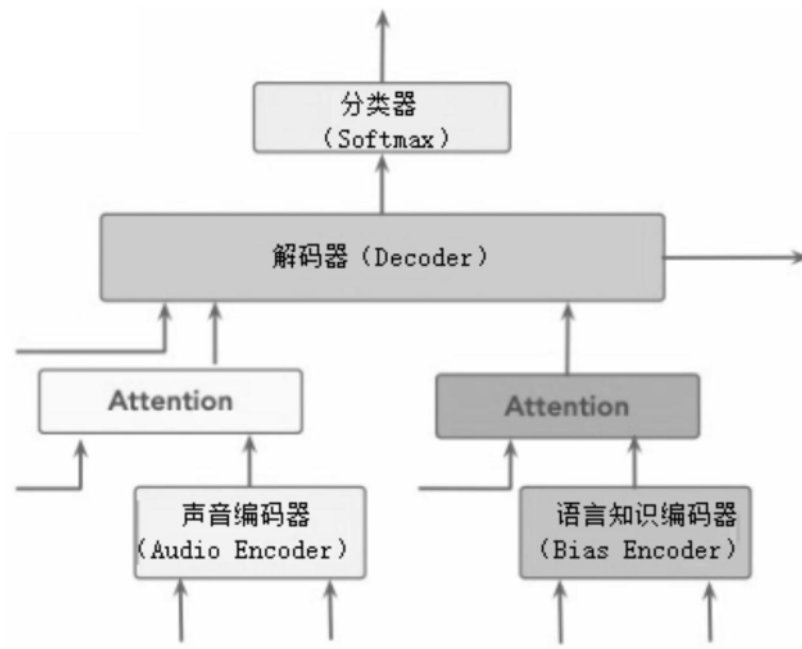


图4

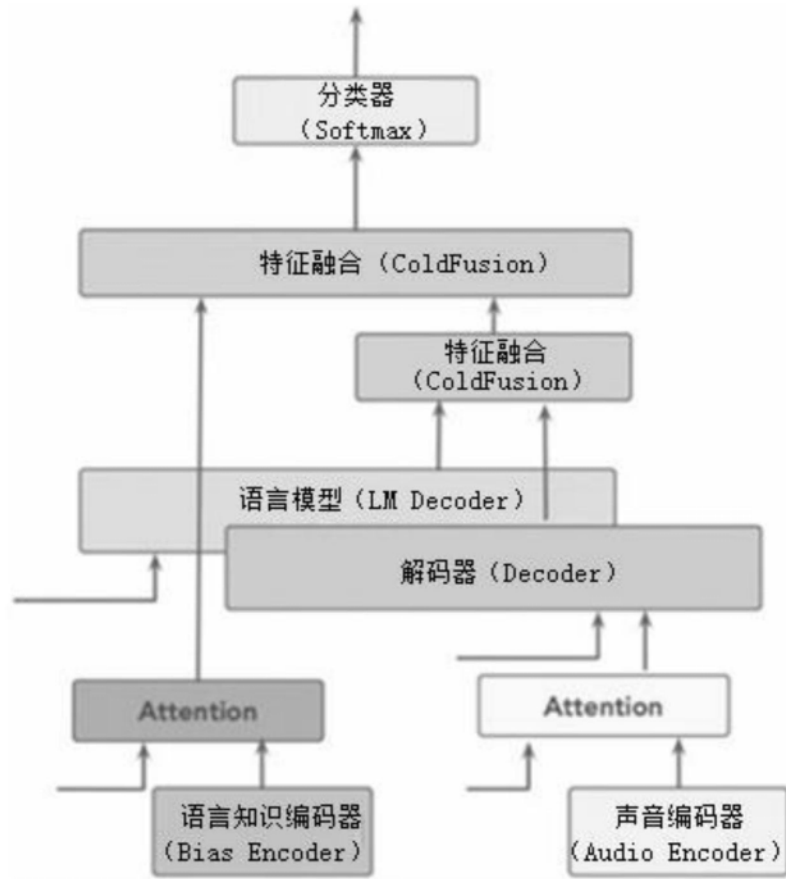


图5

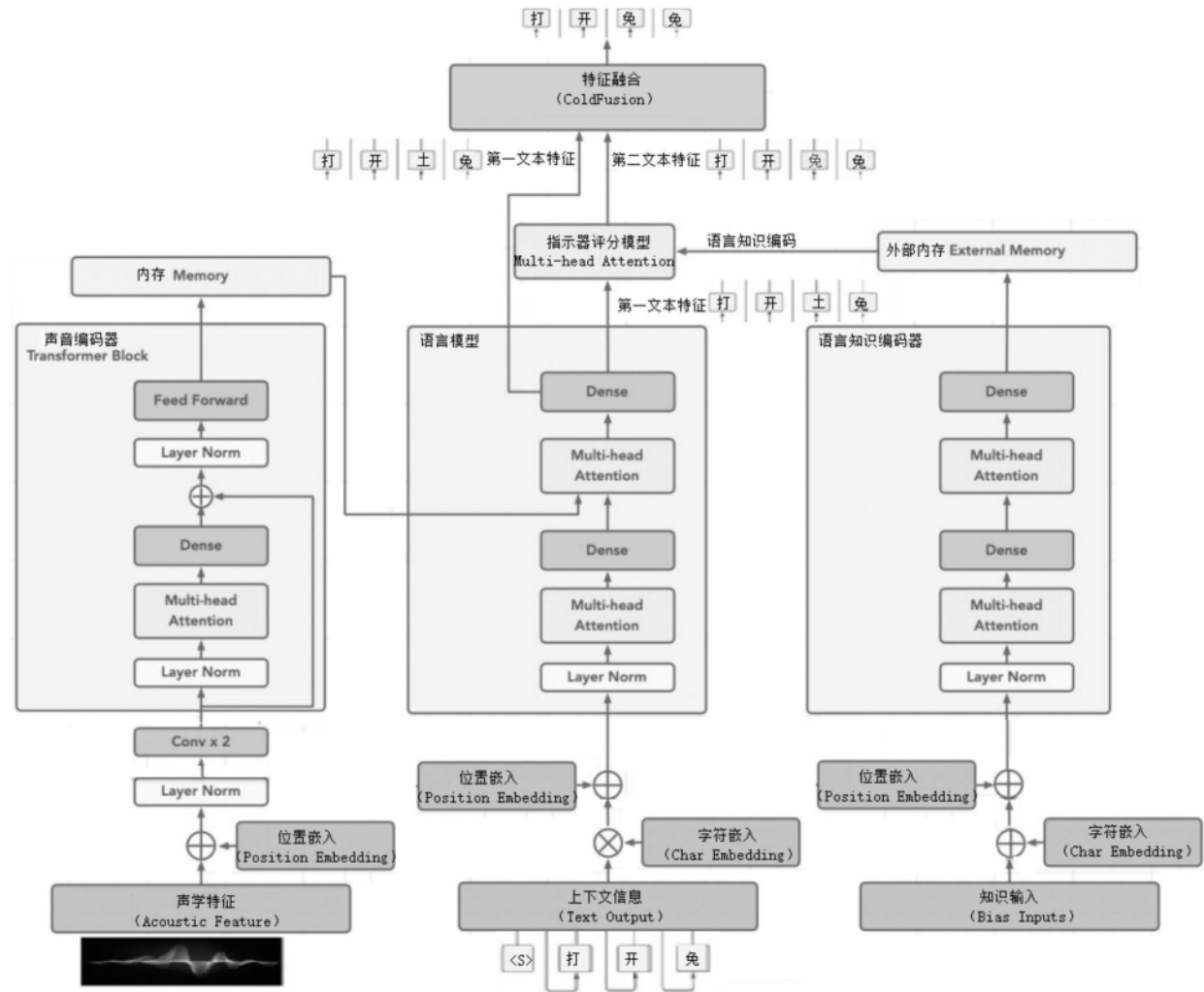


图6

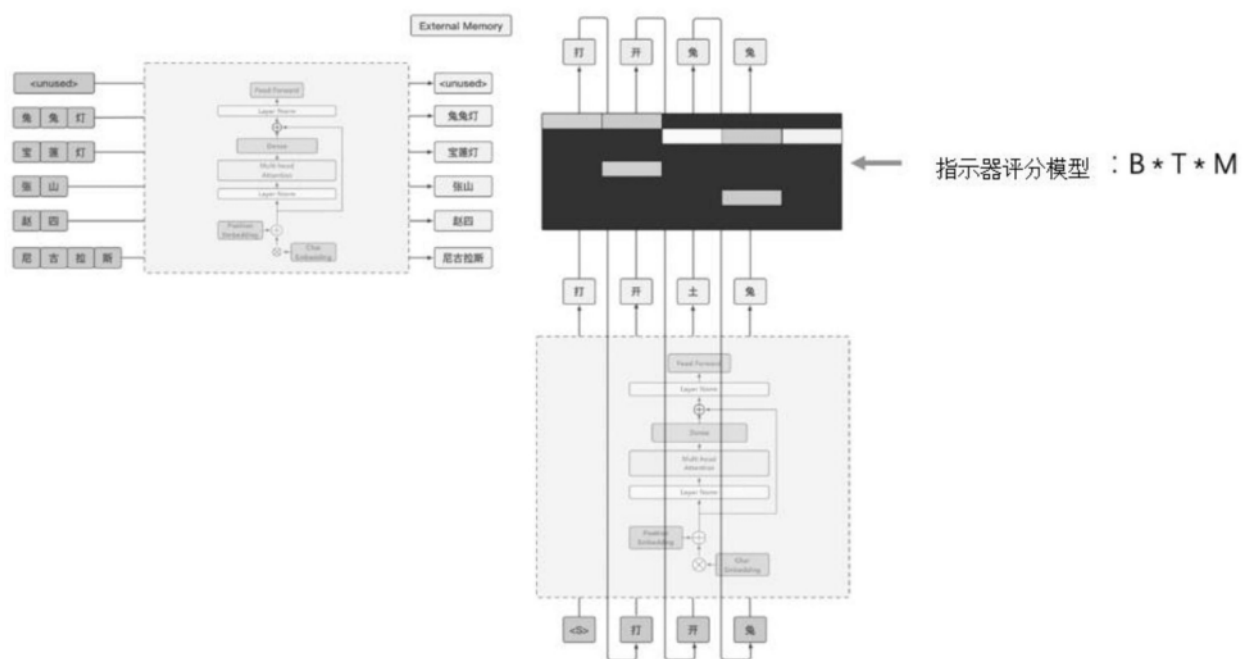


图7