



(12)发明专利申请

(10)申请公布号 CN 111194463 A

(43)申请公布日 2020.05.22

(21)申请号 201880019264.8

G10L 15/19(2013.01)

(22)申请日 2018.08.27

G10L 15/197(2013.01)

(85)PCT国际申请进入国家阶段日  
2019.09.19

G10L 15/183(2013.01)

G10L 15/06(2013.01)

(86)PCT国际申请的申请数据

PCT/CN2018/102544 2018.08.27

(71)申请人 北京嘀嘀无限科技发展有限公司  
地址 100193 北京市海淀区东北旺路西路8  
号院34号楼

(72)发明人 黄辰

(74)专利代理机构 成都七星天知识产权代理有  
限公司 51253  
代理人 袁春晓

(51)Int.Cl.

G10L 15/22(2006.01)

权利要求书7页 说明书22页 附图7页

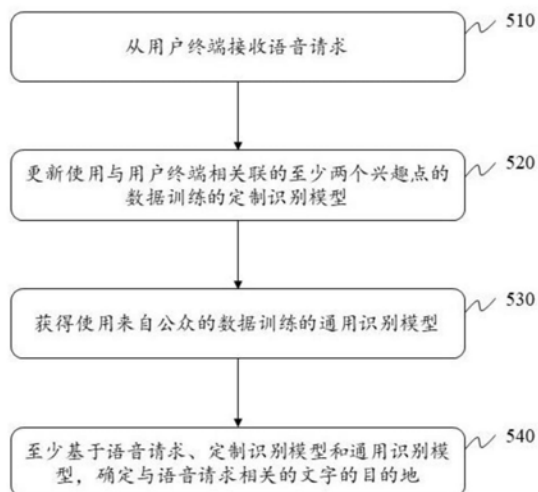
(54)发明名称

用于在移动设备上显示目的地的人工智能系统和方法

(57)摘要

提供了响应于客户终端的语音请求提供线上到线下服务的系统和方法。一种方法包括：从用户终端接收语音请求；响应于语音请求，更新使用与用户终端相关联的至少两个兴趣点的数据训练的定制识别模型；获取使用公众数据训练的通用识别模型；至少基于语音请求、定制识别模型和通用识别模型，确定与语音请求相关联的文字的目的地。

500



1. 一种用于提供线上到线下服务以响应于用户终端的语音请求的一个或以上电子设备的系统,包括:

至少一个存储介质,包括操作系统和与所述操作系统兼容的一组指令,响应于用户终端的语音请求,用于提供线上到线下服务;以及

至少一个处理器,与所述至少一个存储介质通信,其中当执行所述操作系统和所述一组指令时,所述至少一个处理器还用于:

从所述用户终端接收所述语音请求;

获取使用与对应于所述用户终端的至少两个兴趣点相关联的数据训练的定制识别模型;

获取使用公众数据训练的通用识别模型;

至少基于所述语音请求、所述定制识别模型和所述通用识别模型,确定与所述语音请求相关联的文字的目的地;

响应于确定所述文字的目的地,生成包括所述文字的目的地电子信号;以及

将所述电子信号发送到所述用户终端。

2. 根据权利要求1所述的系统,所述至少一个处理器还用于:

响应于所述语音请求,更新所述定制识别模型。

3. 根据权利要求1或2所述的系统,其特征在于,所述至少两个兴趣点包括与所述用户终端的用户相关联的至少两个历史线上到线下服务的至少两个历史出发点和至少两个历史目的地。

4. 根据权利要求1-3任一项所述的系统,其特征在于,未来确定所述文字的目的地,所述至少一个处理器还用于:

基于所述语音请求和所述定制识别模型,确定至少一个定制结果,所述至少一个定制结果中的每一个结果包括定制文字序列和序列概率,所述序列概率表示所述语音请求与所述定制文字序列相关联的概率;

基于所述语音请求和所述通用识别模型确定至少一个通用结果,所述至少一个通用结果,包括通用文字序列和序列概率,所述序列概率表示所述语音请求与所述通用文字序列相关联的概率;以及

从所述至少有一个定制文字序列和所述至少一个通用文字序列中确定一个候选文字序列作为所述文字的目的地,其中所述候选文字序列在所述至少一个定制文字序列和所述至少一个通用文字序列中具有最大的序列概率。

5. 根据权利要求1所述的系统,其特征在于,所述定制识别模型是定制语音模型,所述通用识别模型是通用语音模型,所述至少一个处理器还用于:

基于声学模型和所述语音请求,确定所述语音请求的至少两个声学结果;以及

至少基于所述至少两个声学结果、所述定制语音模型和所述通用语音模型,确定所述文字的目的地。

6. 根据权利要求5所述的系统,其特征在于,未来确定所述至少两个声学结果,所述至少一个处理器还用于:

获取所述语音请求的至少两个帧;

对于所述至少两个帧的每个帧,

识别至少一个声学特征,以及  
基于所述声学模型确定声学结果,其中所述至少一个声学特征是所述声学模型的输入。

7. 根据权利要求6所述的系统,其特征在于,每个声学结果包括至少两个候选语音元素和至少两个相应的声学概率,每个声学概率表示语音请求的所述对应帧与所述对应候选语音元素相关联的概率。

8. 根据权利要求1至4任一项所述的系统,其特征在于,获取所述通用识别模型,所述至少一个处理器还用于:

从公众获取至少两个地点的文字信息;以及

使用所述至少两个位置的所述文字信息作为所述初始语音模型的输入来训练初始语音模型,以获取所述通用识别模型,其中所述通用识别模型是通用语音模型。

9. 根据权利要求8所述的系统,其特征在于,来自所述公众的所述至少两个位置的所述文字信息是从互联网爬取的。

10. 根据权利要求8或9所述的系统,其特征在于,所述通用识别模型是线下训练的。

11. 根据权利要求1所述的系统,其特征在于,与所述用户终端相关联的所述至少两个兴趣点包括以下至少一个:所述用户终端所在的至少两个历史名额、与所述用户终端相关联的区域的至少两个位置,或与所述用户终端相关联的至少两个类似用户的至少两个历史位置。

12. 根据权利要求1所述的系统,其特征在于,所述定制识别模型是一种定制语音模型,所述至少两个兴趣点的所述数据包括所述至少两个兴趣点的文字信息。

13. 响应于客户终端的语音请求,用于提供线上到线下服务的方法,在具有至少一个存储介质和至少一个与所述至少一个存储介质通信的处理器的一个或以上电子设备上实现,包括:

从用户终端接收语音请求;

获取使用与所述用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;

获取使用公众数据训练的通用识别模型;

至少基于所述语音请求、所述定制识别模型和所述通用识别模型确定与所述语音请求相关联的文字的目的地;

响应于确定所述文字的目的地,生成包括所述文字的目的地电子信号;以及

将所述电子信号发送到所述用户终端。

14. 根据权利要求13所述的方法进一步包括:

响应于所述语音请求,更新所述定制识别模型。

15. 根据权利要求13或14所述的方法,其特征在于,所述至少两个兴趣点包括与所述用户终端的用户相关联的至少两个历史线上到线下服务的至少两个历史出发点和至少两个历史目的地。

16. 根据权利要求13-15任一项所述的方法,其特征在于,确定所述文字的目的地包括:

基于所述语音请求和所述定制识别模型确定至少一个定制结果,所述至少一个定制结果中的每一个包括定制文字序列和序列概率,所述序列概率表示所述语音请求与所述定制文字序列相关联的概率;

基于所述语音请求和所述通用识别模型确定至少一个通用结果,所述至少一个通用结果中的每一个包括通用文字序列和序列概率,所述序列概率表示所述语音请求与所述通用文字序列相关联的概率;以及

从所述至少一个定制文字序列和所述至少一个通用文字序列确定一个候选文字序列作为所述文字的目的地,其中所述候选文字序列在所述至少一个定制文字序列和所述至少一个通用文字序列中具有最大的序列概率。

17. 根据权利要求13所述的方法,其特征在于,所述定制识别模型是一种定制语音模型,所述通用识别模型是通用语音模型,所述方法还包括:

基于声学模型和所述语音请求确定所述语音请求的至少两个声学结果;以及

至少基于所述至少两个声学结果、所述定制语音模型和所述通用语音模型,确定所述文字的目的地。

18. 根据权利要求17所述的方法,其特征在于,确定所述至少两个声学结果包括:

获取所述语音请求的至少两个帧;

对于所述至少两个帧的每个帧,

识别至少一个声学特征,以及

基于所述声学模型确定声学结果,其中所述至少一个声学特征是所述声学模型的输入。

19. 根据权利要求18所述的方法,其特征在于,每个声学结果包括至少两个候选语音元素和至少两个相应的声学概率,每个声学概率表示语音请求的所述对应帧与所述对应候选语音元素相关联的概率。

20. 根据权利要求13-16任一项所述的方法,其特征在于,获取所述通用识别模型包括:

从所述公众获取至少两个地点的文字信息;以及

使用所述至少两个位置的所述文字信息作为所述初始语音模型的输入来训练初始语音模型,以获取所述通用识别模型,其中所述通用识别模型是通用语音模型。

21. 根据权利要求20所述的方法,其特征在于,来自所述公众的所述至少两个位置的所述文字信息是从互联网爬取的。

22. 根据权利要求20或21所述的方法,其特征在于,所述通用识别模型是线下训练的。

23. 根据权利要求13所述的方法,其特征在于,与所述用户终端相关联的所述至少两个兴趣点包括以下至少一个:所述用户终端所在的至少两个历史地点、与所述用户终端相关联的区域的至少两个位置,或者与所述用户终端相关联的至少两个类似用户的至少两个历史位置。

24. 根据权利要求13所述的方法,其特征在于,所述定制识别模型是定制语音模型,所述至少两个兴趣点的数据包括所述至少两个兴趣点的文字信息。

25. 一种非暂时性计算机可读介质,包括操作系统和与所述操作系统兼容的至少一组指令,响应于客户终端的语音请求,用于提供线上到线下服务,其中,当由一个或以上电子设备的至少一个处理器执行时,所述至少一组指令使所述至少一个处理器:

从用户终端接收语音请求;

获取使用与所述用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;

获取使用公众数据训练的通用识别模型;

至少基于所述语音请求、所述定制识别模型和所述通用识别模型确定与所述语音请求相关联的文字的目的地；

响应于确定所述文字的目的地，生成电子信号，包括所述文字的目的地；以及  
将所述电子信号发送到所述用户终端。

26. 一种用于提供线上到线下服务以响应于客户终端的语音请求的一个或以上电子设备的人工智能系统，包括：

至少一个目标系统的信息交换端口，其中所述目标系统与用户终端相关联，在所述至少一个信息交换端口和所述用户终端之间通过无线通信从所述用户终端接收语音请求；

至少一个存储介质，包括操作系统和与所述操作系统兼容的一组指令，响应于客户终端的语音请求，用于提供线上到线下服务；以及

至少一个处理器与所述至少一个存储介质通信，其中，当执行所述操作系统和所述指令集时，所述至少一个处理器还用于：

从所述用户终端接收所述语音请求；

获取使用与所述用户终端相关联的至少两个兴趣点相关联的数据训练的定制识别模型；

获取使用公众数据训练的通用识别模型；

至少基于所述语音请求、所述定制识别模型和所述通用识别模型确定与所述语音请求相关联的文字的目的地；

响应于确定所述文字的目的地，生成包括所述文字的目的地和触发代码的电子信号，其中所述触发代码是：

以所述用户终端中安装的应用程序可识别的格式，以及

被配置为使所述应用程序在所述用户终端的界面上生成所述文字目的地的呈现；以及  
将所述电子信号发送到所述目标系统的所述至少一个信息交换端口以使所述至少一个信息交换端口将所述电子信号发送到所述用户终端。

27. 根据权利要求26所述的系统，其特征在於，所述至少一个处理器还用于：

响应于所述语音请求，更新所述定制识别模型。

28. 根据权利要求26或27所述的系统，其特征在於，所述至少两个兴趣点包括与所述用户终端的用户相关联的至少两个历史线上到线下服务的至少两个历史出发点和至少两个历史目的地。

29. 根据权利要求26-28任一项所述的系统，其特征在於，确定所述文字的目的地，所述至少一个处理器还用于：

基于所述语音请求和所述定制识别模型确定至少一个定制结果，所述至少一个定制结果中的每一个包括定制文字序列和序列概率，所述序列概率表示所述语音请求与所述定制文字序列相关联的概率；

基于所述语音请求和所述通用识别模型确定至少一个通用结果，所述至少一个通用结果中的每一个包括通用文字序列和序列概率，所述序列概率表示所述语音请求与所述通用文字序列相关联的概率；以及

从所述至少有一个定制文字序列和所述至少一个通用文字序列确定一个候选文字序列作为所述文字的目的地，其中所述候选文字序列在所述至少一个定制文字序列和所述至

少一个通用文字序列中具有最大的序列概率。

30. 根据权利要求26所述的系统,其特征在於,所述定制识别模型是一种定制语音模型,所述通用识别模型是通用语音模型,所述至少一个处理器还用于:

基于声学模型和所述语音请求,确定所述语音请求的至少两个声学结果;以及

至少基于所述至少两个声学结果、所述定制语音模型和所述通用语音模型,确定所述文字的目的地。

31. 根据权利要求30所述的系统,其特征在於,为了确定所述至少两个声学结果,所述至少一个处理器还用于:

获取所述语音请求的至少两个帧

对于所述至少两个帧的每个帧,

识别至少一个声学特征,以及

基于所述声学模型确定声学结果,其中所述至少一个声学特征是所述声学模型的输入。

32. 根据权利要求31所述的系统,其特征在於,每个声学结果包括至少两个候选语音元素和至少两个相应的声学概率,每个声学概率表示所述语音请求的相应帧与所述对应的候选语音元素相关联的概率。

33. 根据权利要求26-29任一项所述的系统,其特征在於,为了获取所述通用识别模型,所述至少一个处理器还用于:

从所述公众获取至少两个地点的文字信息;以及

使用所述至少两个位置的所述文字信息作为所述初始语音模型的输入来训练初始语音模型,以获取所述通用识别模型,其中所述通用识别模型是通用语音模型。

34. 根据权利要求33所述的系统,其特征在於,来自所述公众的所述至少两个位置的所述文字信息是从互联网爬取的。

35. 根据权利要求33或34所述的系统,其特征在於,所述通用识别模型是线下训练的。

36. 根据权利要求26所述的系统,其特征在於,与所述用户终端相关联的所述至少两个兴趣点包括以下至少一个:所述用户终端所在的至少两个历史地点、与所述用户终端相关联的区域的至少两个位置,或者与所述用户终端相关联的至少两个类似用户的至少两个历史位置。

37. 根据权利要求26所述的系统,其特征在於,所述定制识别模型是定制语音模型,所述至少两个兴趣点的所述数据包括所述至少两个兴趣点的文字信息。

38. 一种用于提供线上到线下服务以响应于客户终端的语音请求,的方法,在具有至少一个信息交换端口、至少一个存储介质,以及与所述至少一个存储介质通信的至少一个处理器的一个或以上电子设备上实现,包括:

从用户终端接收语音请求;

获取使用与所述用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;

获取使用公众数据训练的通用识别模型;

至少基于所述语音请求、所述定制识别模型和所述通用识别模型确定与所述语音请求相关联的文字的目的地;

响应于确定所述文字的目的地,产生包括所述文字的目的地和触发代码的电子信号,

其中所述触发代码是：

以所述用户终端中安装的应用程序可识别的格式，以及被配置为使所述应用程序在所述用户终端的界面上生成所述文字目的地的呈现；以及将所述电子信号发送到所述至少一个信息交换端口，以使所述至少一个信息交换端口将所述电子信号发送到所述用户终端。

39. 根据权利要求38所述的系统进一步包括：

响应于所述语音请求，更新所述定制识别模型。

40. 根据权利要求38或39所述的方法，其特征在于，所述至少两个兴趣点包括与所述用户终端的用户相关联的至少两个历史线上到线下服务的至少两个历史出发点和至少两个历史目的地。

41. 根据权利要求38或40所述的方法，其特征在于，确定所述文字的目的地包括：

基于所述语音请求和所述定制识别模型确定至少一个定制结果，所述至少一个定制结果中的每一个包括定制文字序列和序列概率，所述序列概率表示所述语音请求与所述定制文字序列相关联的概率；

基于所述语音请求和所述通用识别模型确定至少一个通用结果，所述至少一个通用结果中的每一个包括通用文字序列和序列概率，所述序列概率表示所述语音请求与所述通用文字序列相关联的概率；以及

从所述至少一个定制文字序列和所述至少一个通用文字序列确定一个候选文字序列作为所述文字的目的地，其中所述候选文字序列在所述至少一个定制文字序列和所述至少一个通用文字序列中具有最大的序列概率。

42. 根据权利要求38所述的方法，其特征在于，所述定制识别模型是一种定制语音模型，所述通用识别模型是通用语音模型，所述方法还包括：

基于声学模型和所述语音请求，确定所述语音请求的至少两个声学结果；以及

至少基于所述至少两个声学结果，所述定制语音模型和所述通用语音模型，确定所述文字的目的地。

43. 根据权利要求42所述的方法，其特征在于，确定所述至少两个声学结果包括：

获取所述语音请求的至少两个帧；

对于所述至少两个帧的每个帧，

识别至少一个声学特征，以及

基于所述声学模型确定声学结果，其中所述至少一个声学特征是所述声学模型的输入。

44. 根据权利要求43所述的方法，其特征在于，每个声学结果包括至少两个候选语音元素和至少两个相应的声学概率，每个声学概率表示语音请求的所述对应帧与所述对应候选语音元素相关联的概率。

45. 根据权利要求38-41任一项所述的方法，其特征在于，所述获取所述通用识别模型包括：

从所述公众获取至少两个地点的文字信息；以及

使用所述至少两个位置的所述文字信息作为所述初始语音模型的输入来训练初始语音模型，以获取所述通用识别模型，其中所述通用识别模型是通用语音模型。

46. 根据权利要求45所述的方法,其特征在於,来自所述公众的所述至少两个位置的所述文字信息是从互联网爬取的。

47. 根据权利要求45或46所述的方法,其特征在於,所述通用识别模型是线下训练的。

48. 根据权利要求38所述的方法,其特征在於,与所述用户终端相关联的所述至少两个兴趣点包括以下至少一个:所述用户终端所在的至少两个历史地点、与所述用户终端相关联的区域的至少两个位置,或者与所述用户终端相关联的至少两个类似用户的至少两个历史位置。

49. 根据权利要求38所述的方法,其特征在於,所述定制识别模型是定制语音模型,所述至少两个兴趣点的所述数据包括所述至少两个兴趣点的文字信息。

50. 一种非暂时性计算机可读介质,包括操作系统和与所述操作系统兼容的至少一组指令,响应于客户终端的语音请求,用于提供线上到线下服务,其中,当由一个或以上电子设备的至少一个处理器执行时,所述至少一组指令使所述至少一个处理器:

从用户终端接收语音请求;

获取使用与所述用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;

获取使用公众数据训练的通用识别模型;

至少基于所述语音请求、所述定制识别模型和所述通用识别模型,确定与所述语音请求相关联的文字的目的地;

响应于确定所述文字的目的地,生成包括所述文字的目的地和触发代码的电子信号,其中所述触发代码是:

以所述用户终端中安装的应用程序可识别的格式,以及

被配置为使所述应用程序在所述用户终端的界面上生成所述文字目的地的呈现;以及

将所述电子信号发送到所述一个或以上电子设备的至少一个信息交换端口以使所述至少一个信息交换端口以将所述电子信号发送到所述用户终端。

51. 一种用于提供线上到线下服务以响应于客户终端的语音请求的人工智能系统,包括:

请求获取模块,被配置为从用户终端接收语音请求;

模型获取模块,被配置用于获取使用与所述用户终端相关联的至少两个兴趣点的数据训练的定制识别模型和使用来自公众的数据训练的通用识别模型;

目的地确定模块,被配置为至少基于所述语音请求、所述定制识别模型和所述通用识别模型确定与所述语音请求相关联的文字的目的地。

## 用于在移动设备上显示目的地的人工智能系统和方法

### 技术领域

[0001] 本申请一般涉及使用人工智能在线上到线下服务系统中确定并在用户终端设备显示文字的目的地系统和方法以响应于用户移动设备的语音请求。

### 背景技术

[0002] 在现代社会中,线上到线下服务的应用正变得越来越普遍。然而,当用户请求服务时,用户必须在安装在用户移动终端中的应用程序中手动拼写或输入信息。用户可能不熟悉用户拼写或输入的信息(例如,服务递送目的地、服务开始起始位置、服务项目等)。例如,用户仅知道信息的发音,但不知道拼写或其写入。拼写或书面形式经常导致错误的信息并且花费很长时间。通过人工智能(AI)使用训练有素的语音识别模型,用户可以通过语音请求服务,而不是手动拼写或输入信息。

[0003] 然而,现有技术的技术解决方案通常线下训练语音识别模型,并定期更新训练数据。现有技术的问题是语音识别模型没有实时训练,这导致不准确的语音识别结果。因此,希望提供用于在请求线上到线下服务时准确识别语音片段中的位置的系统和方法。

### 发明内容

[0004] 本申请的一个方面是介绍一种用于提供线上到线下服务以响应于用户终端的语音请求的一个或以上电子设备的系统,包括:至少一个存储介质,包括响应于用户终端的语音请求用于提供线上到线下服务的操作系统和与操作系统兼容的一组指令;和与至少一个存储介质通信的至少一个处理器,其中,当执行操作系统和所述一组指令时,至少一个处理器还用于:接收用户终端的语音请求;获取与至少两个用户终端的兴趣点相关联的使用数据训练的定制识别模型;获取使用公众数据训练的通用识别模型;至少基于语音请求,定制识别模型和通用识别模型确定与语音请求相关联的文字的目的地;响应于确定文字的目的地,生成包括文字的目的地电子信号;并将电子信号发送给用户终端。

[0005] 在一些实施例中,所述至少一个处理器还用于:响应于语音请求,更新定制识别模型。

[0006] 在一些实施例中,至少两个兴趣点包括至少两个与用户终端的用户相关联的历史线上到线下服务的至少两个历史出发点和至少两个历史目的地。

[0007] 在一些实施例中,为了确定文字的目的地,至少一个处理器还用于:基于语音请求和定制识别模型,确定至少一个定制结果,所述至少一个定制结果中的每一个包括定制文字序列和序列概率,所述序列概率表示语音请求与定制文字序列相关联的概率;基于语音请求和通用识别模型确定至少一个通用结果,所述至少一个通用结果中的每一个包括通用文字序列和序列概率,所述序列概率表示语音请求与通用文字序列相关联的概率;从所述至少一个定制文字序列和所述至少一个通用文字序列确定一个候选文字序列作为文字的目的地,其中候选文字序列在至少一个定制文字序列和至少一个通用文字序列中具有最大的序列概率。

[0008] 在一些实施例中,定制识别模型是定制语音模型,通用识别模型是通用语音模型,并且该至少一个处理器还用于:基于声学模型和语音请求确定语音请求的至少两个声学结果;并基于至少两个声学结果、定制语音模型和通用语音模型确定文字的目的地。

[0009] 在一些实施例中,为了确定至少两个声学结果,所述至少一个处理器还用于:获取语音请求的至少两个帧;基于声学模型,对于所述的至少两个帧的每个帧,识别至少一个声学特征,并确定声学结果,其中所述至少一个声学特征是声学模型的输入。

[0010] 在一些实施例中,每个声学结果包括至少两个候选语音元素和至少两个对应的声学概率,每个声学概率表示语音请求的相应帧与对应的候选语音元素相关联的概率。

[0011] 在一些实施例中,为了获取通用识别模型,所述至少一个处理器还用于:从公众获取至少两个位置的文字信息;使用至少两个位置的文字信息作为初始语音模型的输入来训练初始语音模型以获取通用识别模型,其中通用识别模型是通用语音模型。

[0012] 在一些实施例中,来自公众的至少两个位置的文字信息是从互联网爬取的。

[0013] 在一些实施例中,通用识别模型是线下训练的。

[0014] 在一些实施例中,与用户终端相关联的至少两个兴趣点包括以下中的至少一个:用户终端曾经位于的至少两个历史位置,与用户终端相关联的区域的至少两个位置,或者与用户终端相关联的至少两个类似用户的至少两个历史位置。

[0015] 在一些实施例中,定制识别模型是定制语音模型,并且至少两个兴趣点的数据包括至少两个兴趣点的文字信息。

[0016] 根据本申请的另一方面,一种用于提供线上到线下服务以响应于用户终端的语音请求,的方法可以在具有至少一个存储介质和至少一个与所述至少一个存储介质通信的处理器的一个或以上电子设备上实现,所述方法包括:从用户终端接收语音请求;获取使用与用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;获取使用公众数据训练的通用识别模型;至少基于语音请求、定制识别模型和通用识别模型,确定与语音请求相关联的文字的目的地;响应于确定文字的目的地,产生包括文字的目的地电子信号;并将电子信号发送给用户终端。

[0017] 在一些实施例中,该方法还包括:响应于语音请求,更新定制识别模型。

[0018] 在一些实施例中,至少两个兴趣点包括与用户终端的用户相关联的至少两个历史上到线下服务的至少两个历史出发点和至少两个历史目的地。

[0019] 在一些实施例中,确定文字的目的地包括:基于语音请求和定制识别模型确定至少一个定制结果,所述至少一个定制结果中的每一个包括定制文字序列和序列概率,所述序列概率表示语音请求与定制文字序列相关联的概率;根据语音请求和通用识别模型确定至少一个通用结果,所述至少一个通用结果中的每一个包括通用文字序列和序列概率,所述序列概率表示语音请求与通用文字序列相关联的概率;从所述至少一个定制文字序列和所述至少一个通用文字序列确定一个候选文字序列作为文字的目的地,其中候选文字序列在至少一个定制文字序列和至少一个通用文字序列中具有最大的序列概率。

[0020] 在一些实施例中,定制识别模型是定制语音模型,通用识别模型是通用语音模型,该方法还包括:基于声学模型和语音请求确定语音请求的至少两个声学结果;并基于至少两个声学结果、定制语音模型和通用语音模型来确定文字的目的地。

[0021] 在一些实施例中,确定至少两个声学结果包括:获取语音请求的至少两个帧;对于

至少两个帧的每个帧,基于声学模型,识别至少一个声学特征,并确定声学结果,其中,所述至少一个声学特征是声学模型的输入。

[0022] 在一些实施例中,每个声学结果包括至少两个候选语音元素和至少两个对应的声学概率,每个声学概率表示语音请求的相应帧与对应的候选语音元素相关联的概率。

[0023] 在一些实施例中,获取通用识别模型包括:从公众获取至少两个位置的文字信息;使用至少两个位置的文字信息作为初始语音模型的输入来训练初始语音模型以获取通用识别模型,其中通用识别模型是通用语音模型。

[0024] 在一些实施例中,来自公众的至少两个位置的文字信息是从互联网爬取的。

[0025] 在一些实施例中,通用识别模型是线下训练的。

[0026] 在一些实施例中,与用户终端相关联的至少两个兴趣点包括以下中的至少一个:用户终端曾经定位的至少两个历史位置,与用户终端相关联的区域的至少两个位置,或者与用户终端相关联的至少两个类似用户的至少两个历史位置。

[0027] 在一些实施例中,定制识别模型是定制语音模型,并且至少两个兴趣点的数据包括至少两个兴趣点的文字信息。

[0028] 根据本申请的另一方面,一种非暂时性计算机可读介质,包括操作系统和与操作系统兼容的至少一组指令,用于从用户终端提供线上到线下服务响应,以及语音请求,其中,当由一个或以上电子设备的至少一个处理器执行时,所述至少一组指令使所述至少一个处理器:从用户终端接收语音请求;获取使用与用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;获取使用公众数据训练的通用识别模型;至少基于语音请求、定制识别模型和通用识别模型,确定与语音请求相关联的文字的目的地;响应于确定文字的目的地,生成包括文字的目的地电子信号;并将电子信号发送给用户终端。

[0029] 根据本申请的又一方面,一种用于提供线上到线下服务以响应于用户终端的语音请求的一个或以上电子设备的人工智能系统,包括:至少一个目标系统的信息交换端口,其中,目标系统与用户终端相关联,以通过至少一个信息交换端口和用户终端之间的无线通信从用户终端接收语音请求;至少一个存储介质,包括操作系统和与操作系统兼容的一组指令,响应于用户终端的语音请求,用于提供线上到线下服务;至少一个处理器与至少一个存储介质通信,其中当执行操作系统和所述指令集时,至少一个处理器还用于:从用户终端接收语音请求;获取使用与用户终端相关联的至少两个兴趣点相关联的数据训练的定制识别模型;获取使用公众数据训练的通用识别模型;至少基于语音请求、定制识别模型和通用识别模型,确定与语音请求相关联的文字的目的地;响应于确定文字的目的地,生成包括文字的目的地和触发代码的电子信号,其中触发代码是:以安装在用户终端中的应用程序可识别的,并且被配置为使应用程序在用户终端的界面上生成文字的目的地格式呈现;并将所述电子信号发送至所述目标系统的至少一个信息交换端口,以使所述至少一个信息交换端口将所述电子信号发送至所述用户终端。

[0030] 根据本申请的另一方面,一种用于提供线上到线下服务以响应于用户终端的语音请求的方法可以在具有至少一个信息交换端口、至少一个存储介质、以及与至少一个存储介质通信的至少一个处理器的一个或以上电子设备上实现,所述方法包括:从用户终端接收语音请求;获取使用与用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;获取使用公众数据训练的通用识别模型;至少基于语音请求、定制识别模型和通用识别模

型,确定与语音请求相关联的文字的目的地;响应于确定文字的目的地,生成包括文字的目的地和触发代码的电子信号,其中触发代码是:以安装在用户终端中的应用程序可识别的,并且被配置为使应用程序在用户终端的界面上生成文字的目的地格式呈现;并将所述电子信号发送至所述至少一个信息交换端口,以使所述至少一个信息交换端口将所述电子信号发送至所述用户终端。

[0031] 根据本申请的又一方面,一种非暂时性计算机可读介质,包括操作系统和与操作系统兼容的至少一组指令用于提供线上到线下服务以响应于客户终端的语音请求,其中,当由一个或以上电子设备的至少一个处理器执行时,所述至少一组指令使所述至少一个处理器:从用户终端接收语音请求;获取使用与用户终端相关联的至少两个兴趣点的数据训练的定制识别模型;获取使用公众数据训练的通用识别模型;至少基于语音请求、定制识别模型和通用识别模型,确定与语音请求相关联的文字的目的地;响应于确定文字的目的地,生成包括文字的目的地和触发代码的电子信号,其中触发代码是:以安装在用户终端中的应用程序可识别的,并且被配置为使应用程序在用户终端的界面上生成文字的目的地格式呈现;并将电子信号发送到一个或以上电子设备的至少一个信息交换端口,以使至少一个信息交换端口将电子信号发送给用户终端。

[0032] 根据本申请的又一方面,一种用于提供线上到线下服务以响应于用户终端的语音请求的人工智能系统,包括:被配置为从用户终端接收语音请求的请求获取模块;被配置为用于获取使用与用户终端相关联的至少两个兴趣点的数据训练的定制识别模型和使用来自公众的数据训练的通用识别模型的模型获取模块;被配置为至少基于语音请求,定制识别模型和通用识别模型确定与语音请求相关联的文字的目的地的目标确定模块。

[0033] 本申请的一部分附加特性可以在下面的描述中进行说明。通过对以下描述和相应附图的研究或者对实施例的生产或操作的了解,本申请的一部分附加特性对于本领域技术人员是明显的。本申请的特征可以通过对以下描述的具体实施例的各种方面的方法、手段和组合的实践或使用得以实现和达到。

## 附图说明

[0034] 本申请将通过示例性实施例进行进一步描述。这些示例性实施例将通过附图进行详细描述。这些实施例是非限制性的示例性实施例,在这些实施例中,各图中相同的编号表示相似的结构,其中:

[0035] 图1是根据本申请的一些实施例所示的示例性人工智能(AI)系统的示意图;

[0036] 图2是根据本申请的一些实施例所示的计算设备的示例性硬件和/或软件组件的示意图;

[0037] 图3是根据本申请的一些实施例所示的移动设备的示例性硬件组件和/或软件组件的示意图;

[0038] 图4是根据本申请的一些实施例所示的示例性处理引擎的框图;

[0039] 图5A是根据本申请的一些实施例所示的确定文字的目的地示例性过程的流程图;

[0040] 图5B是根据本申请的一些实施例所示的用于呈现文字的目的地用户终端的示例性用户界面;

[0041] 图6A是根据本申请的一些实施例所示的将语音请求识别到文字的的目的地的示例性过程的示意图；

[0042] 图6B是根据本申请的一些实施例所示的将语音请求识别到文字的的目的地的示例性过程的示意图；

[0043] 图7是根据本申请的一些实施例所示的用于确定文字的的目的地的示例性过程的流程图；

[0044] 图8是根据本申请的一些实施例所示的用于确定文字的的目的地的示例性过程的流程图；

[0045] 图9是根据本申请的一些实施例所示的用于确定至少两个声学结果的示例性过程的流程图；

[0046] 图10是根据本申请的一些实施例所示的用于确定文字的的目的地的示例性过程的示意图；以及

[0047] 图11是根据本申请的一些实施例所示的用于训练通用识别模型的示例性过程的流程图。

### 具体实施方式

[0048] 以下描述是为了使本领域的普通技术人员能够实施和利用本申请,并且该描述是在特定的应用场景及其要求的环境下提供的。对于本领域的普通技术人员来讲,显然可以对所披露的实施例作出各种改变,并且在不偏离本申请的原则和范围的情况下,本申请中所定义的普遍原则可以适用于其他实施例和应用场景。因此,本申请并不限于所描述的实施例,而应该被给予与权利要求一致的最广泛的范围。

[0049] 本申请中所使用的术语仅用于描述特定的示例性实施例,并不限制本申请的范围。如本申请使用的单数形式“一”、“一个”及“该”可以同样包括复数形式,除非上下文明确提示例外情形。还应当理解,如在本申请说明书中,术语“包括”和/或“包含”仅提示存在所述特征、整体、步骤、操作、组件和/或部件,但并不排除存在或添加一个或以上其他特征、整体、步骤、操作、组件、部件和/或其组合的情况。

[0050] 在考虑了作为本申请一部分的附图的描述内容后,本申请的特征和特点以及操作方法、结构的相关元素的功能、各部分的组合、制造的经济性变得显而易见。然而,应当理解,附图仅仅是为了说明和描述的目的,并不旨在限制本申请的范围。应当理解的是,附图并不是按比例绘制的。

[0051] 本申请中使用了流程图用来说明根据本申请的一些实施例的系统所执行的操作。应当理解的是,流程图中的操作可以不按顺序执行。相反,可以按照倒序或同时处理各种步骤。同时,也可以将一个或以上其他操作添加到这些流程图中。也可以从流程图中删除一个或以上操作。

[0052] 本申请的一个方面涉及响应于用户终端的语音请求用于提供线上到线下服务的系统和方法,然后在目标移动设备上显示与语音相关联的目的地,例如发送语音的智能手机。为此,系统和方法可以通过他/她的移动电话实时记录用户的活动,并且更新对于实时训练语音识别模型有用的目标活动数据(例如,与某个位置有关的语音或文本数据)。然后,系统可以使用目标活动数据作为训练数据来实时训练语音识别模型。由于训练数据是用户

特定的,因此语音识别模型是专门为用户定制模型。训练数据可以包括与用户终端相关联的至少两个历史兴趣点的文字信息(例如,通过文本、代码或系统可识别的任何书面形式表达的信息和/或可转换为人类用户可识别的书面形式的信息)。最后,为了确定语音所说的内容,系统和方法可以通过将定制识别模型的结果与通用识别模型的结果进行比较来确定从语音请求识别的文字的目的地,并显示文字的目的地让用户确认。其系统和方法可以提高语音识别的准确性并改善用户体验。

[0053] 图1是根据本申请的一些实施例的示例性线上到线下服务系统100的示意图。例如,线上线下服务AI系统100可以是用于运输服务的线上到线下服务平台,例如出租车、司机服务、运送车辆、拼车、公共汽车服务、司机租用、班车服务和在线导航服务。线上线下服务AI系统100可以是在线平台,包括服务器110、网络120、用户终端130和存储器140。该服务器110可包含处理引擎112。

[0054] 服务器110可以被配置为与语音请求有关的处理信息和/或数据。例如,服务器110可以训练定制识别模型和/或通用识别模型。又例如,服务器110可以使用定制识别模型和/或通用识别模型从语音请求中识别文字的目的地。文字的目的地可以通过文本、代码或智能设备可识别的任何书面形式和/或可转换为人类用户可识别的书面形式表达的地方。仅用于说明目的,当前本申请使用文本作为文字的目的地。在一些实施例中,服务器110可以是单个服务器,也可以是服务器组。所述服务器组可以是集中式的,也可以是分布式的(例如,服务器110可以是分布式的系统)。在一些实施例中,服务器110可以是本地的,也可以是远程的。例如,服务器110可以通过网络120访问存储于用户终端130或存储器140中的信息和/或数据。又例如,服务器110可以连接用户终端130和/或存储器140以访问存储的信息和/或数据。在一些实施例中,服务器110可以在云平台上实施。仅作为示例,该云平台可以是私有云、公共云、混合云、社区云、分布云、内部云、多层云等或其任意组合。在一些实施例中,服务器110可以在本申请中的图2描述的包含了一个或以上组件的计算设备200上执行。

[0055] 在一些实施例中,服务器110可以包括处理引擎112。处理引擎112可以处理与语音请求有关的信息和/或数据,以执行本申请中描述的一个或以上的功能。例如,处理引擎112可以训练定制识别模型和/或通用识别模型。又如例,处理引擎112可以使用定制识别模型和/或通用识别模型从语音请求识别文字的目的地。在一些实施例中,所述处理引擎112可包括一个或以上处理引擎(例如,单芯片处理引擎或多芯片处理引擎)。仅作为示例,处理引擎112可以包括一个或以上硬件处理器,例如中央处理单元(CPU)、特定应用集成电路(ASIC)、特定应用指令集处理器(ASIP)、图像处理单元(GPU)、物理运算处理单元(PPU)、数字信号处理器(DSP)、现场可编程门阵列(FPGA)、可编程逻辑设备(PLD)、控制器、微控制器单元、精简指令集计算机(RISC)、微处理器等或其任意组合。

[0056] 网络120可以促进信息和/或数据的交换。在一些实施例中,线上线下服务AI系统100的一个或以上组件(例如,服务器110、用户终端130和存储器140)可以通过网络120将信息和/或数据发送到线上线下服务AI系统100中的其他组件。例如,服务器110可以通过网络120从用户终端130接收语音请求。在一些实施例中,网络120可以为任意形式的有线或无线网络,或其任意组合。仅作为示例,网络130可以是缆线网络、有线网络、光纤网络、远程通信网络、内部网络、互联网、局域网络(LAN)、广域网络(WAN)、无线局域网络(WLAN)、城域网

(MAN)、公共交换电话网络(PSTN)、蓝牙网络、紫蜂网络、近场通讯(NFC)网络等或其任意组合。在一些实施例中,网络120可以包括一个或以上网络接入点。例如,网络120可以包括有线或无线网络接入点,例如基站和/或互联网交换点120-1、120-2……,线上线下服务AI系统100的一个或以上组件可以通过所述接入点连接到网络120以在它们之间交换数据和/或信息。

[0057] 用户终端130可以是用户用来请求线上到线下服务的移动设备,例如出租车。用户终端130也可以是用户用来请求线上线下服务的其他类型的电子设备。例如,用户可以使用用户终端130向服务器110发送语音请求以确定语音请求的文字的目的地。又例如,用户终端130可以从服务器110接收包括文字的目的地电子信号,并在用户终端130的界面上显示文字的目的地。在一些实施例中,用户终端130可以是移动设备130-1、平板电脑130-2、膝上型电脑130-3、机动车辆130-4中的内置设备等,或其任何组合。在一些实施例中,移动设备130-1可以是可穿戴设备、智能移动设备、虚拟现实设备、增强现实设备等,或其任何组合。在一些实施例中,该可穿戴设备可以是智能手镯、智能鞋袜、智能眼镜、智能头盔、智能手表、智能衣服、智能背包、智能配件等,或其任意组合。在一些实施例中,智能移动设备可以是智能电话、个人数字助理(PDA)、游戏设备、导航设备、销售点(POS)等,或其任意组合。在一些实施例中,虚拟现实设备和/或增强型虚拟现实设备可以是虚拟现实头盔、虚拟现实眼镜、虚拟现实眼罩、增强现实头盔、增强现实眼镜、增强现实眼罩等,或其任意组合。例如,虚拟现实设备和/或增强现实设备可以是Google Glass™、RiftCon™、Fragments™、Gear VR™等。在一些实施例中,机动车辆的内置设备130-4可以是车载计算机、车载电视等。

[0058] 在一些实施例中,用户终端130可以是具有定位技术的设备,用于定位用户和/或用户终端130的位置。本申请中使用的定位技术可以包括全球定位系统(GPS)、全球卫星导航系统(GLONASS)、北斗导航系统(COMPASS)、伽利略定位系统、准天顶卫星系统(QZSS)、无线保真(Wi-Fi)定位技术等,或其任意组合。以上定位技术中的一个或以上可以在本申请中交换使用。在一些实施例中,用户终端130还可以包括至少一个网络端口。该至少一个网络端口可以被配置为通过网络120向AI系统100中的一个或以上组件(例如,服务器110、存储器140)发送信息和/或从其接收信息。在一些实施例中,用户终端130可以在具有一个或以上图2中所示的组件的计算设备200上实现,或者在本申请中具有一个或以上图3中所示的组件的移动设备300上实现。

[0059] 存储器140可以储存数据和/或指令。例如,存储器140可以储存从用户终端130获取的数据。又如此,存储器140可以储存服务器110执行或用于执行本申请中描述的示例性方法的数据和/或指令。在一些实施例中,存储器140可以是大容量存储器、可移动存储器、易失性读写存储器、只读存储器(ROM)等,或其任意组合。示例性的大容量存储器可以包括磁盘、光盘、固态硬盘等。示例性可移动存储器可以包括闪存驱动器、软盘、光盘、存储卡、压缩盘、磁带等。示例性易失性读写存储器可以包括随机存取存储器(RAM)。示例性RAM可包括动态随机存取存储器(DRAM)、双倍数据速率同步动态随机存取存储器(DDR SDRAM)、静态随机存取存储器(SRAM)、晶闸管随机存取存储器(T-RAM)和零电容随机存取存储器(Z-RAM)等。示例性只读存储器可以包括掩模型只读存储器(MROM)、可编程只读存储器(PROM)、可擦除可编程只读存储器(EPROM)、电可擦除可编程只读存储器(EEPROM)、光盘只读存储器(CD-ROM)和数字多功能磁盘只读存储器等。在一些实施例中,所述存储器140可在云平台上实

现。仅作为示例,该云平台可以是私有云、公共云、混合云、社区云、分布云、内部云、多层云等,或其任意组合。

[0060] 在一些实施例中,存储器140可以包括至少一个网络端口,与AI系统100中的其他设备通信。例如,存储器140可以连接到网络120,通过至少一个网络端口与线上线下服务AI系统100的一个或以上组件(例如,服务器110、用户终端130)通信。线上线下服务AI系统100中的一个或以上组件可以通过网络120访问存储在存储器140中的数据或指令。在一些实施例中,存储器140可以与线上线下服务AI系统100中的一个或以上组件(例如,服务器110、用户终端130)直接连接或通信。在一些实施例中,存储器140可以是服务器110的一部分。

[0061] 在一些实施例中,线上线下服务AI系统100的一个或以上组件(例如,服务器110、用户终端130)可以访问存储器140。在一些实施例中,当满足一个或以上条件时,线上线下服务AI系统100的一个或以上组件可以读取和/或修改与用户和/或公众有关的信息。例如,服务器110可以在完成服务之后读取和/或修改一个或以上用户的信息。

[0062] 在一些实施例中,线上线下服务AI系统100的一个或以上组件(例如,服务器110、用户终端130和存储器140)可以通过有线和/或无线通信以电子和/或电磁信号的形式彼此通信。在一些实施例中,AI系统100还可以包括至少一个信息交换端口。至少一个交换端口可以被配置为用于在AI系统100中的任何电子设备之间接收和/或发送与服务请求有关的信息(例如,以电子信号和/或电磁信号的形式)。例如,至少一个信息交换端口可以通过服务器110和用户终端130之间的无线通信从用户终端130接收语音请求。又例如,所述至少一个信息交换端口可以通过无线通信向用户终端130发送包括确定的文字的目的地的电磁信号。在一些实施例中,所述至少一个信息交换端口可以是天线、网络接口、网络端口等的一个或以上,或其任何组合。例如,至少一个信息交换端口可以是连接到服务器110以向其发送信息和/或接收从其发送的信息的网络端口。

[0063] 图2是根据本申请的一些实施例所示的计算设备200的示例性硬件和软件组件的示意图,在该计算设备200上可以实现服务器110和/或用户终端130。例如,所述处理引擎112可以在所述计算设备200上实施并执行本申请所披露的所述处理引擎112的功能。

[0064] 计算设备200可用于实现本申请的AI系统100。计算设备200可用于实现AI系统100的任何组件,其执行本申请中披露的一个或以上功能。例如,处理引擎112可以在计算设备200上通过其硬件、软件程序、固件或其组合实现。尽管仅示出了一个这样的计算机,但是为了方便,与本文所述的线上线下服务有关的计算机功能可以在多个类似平台上以分布式方式实现,以分配处理负荷。

[0065] 例如,计算设备200可以包括与网络相连接通信(COM)端口250,以实现数据通信。COM端口250可以是任何网络端口或信息交换端口,以便于数据通信。计算设备200还可以包括处理器(例如,处理器220),其形式为一个或以上处理器(例如,逻辑电路),用于执行程序指令。例如,处理器包括其中的接口电路和处理电路。接口电路可以被配置为从总线210接收电信号,其中电信号编码用于处理电路的结构化数据和/或指令。处理电路可以进行逻辑计算,然后将结论、结果和/或指令编码确定为电信号。处理电路还可以生成包括结论或结果的电子信号(例如,文字的目的地)和触发代码。在一些实施例中,触发代码可以由AI系统100中的电子设备(例如,用户终端130)的操作系统(或其中安装的应用程序)可识别的格式。例如,触发代码可以是指令、代码、标记、符号等,或其任何组合,其可以激活移动电话的

某些功能和/或操作,或者让移动电话执行预定的程序。在一些实施例中,触发代码可以被配置用于电子设备的操作系统(或应用程序),以在电子设备的接口上生成结论或结果的呈现(例如,文字的目的地)。然后,接口电路可以通过总线210从处理电路发出电信号。

[0066] 示例性计算设备可以包括内部通信总线210、程序存储和不同形式的数据存储,包括:例如,磁盘270,以及只读存储器(ROM) 230或随机存取存储器(RAM) 240,用于由计算设备处理和/或发送的各种数据文件。示例性计算设备也可以包括储存于ROM230、RAM240和/或其他形式的非暂时性存储介质中的能够被处理器220执行的程序指令。本申请的方法和/或流程可以以程序指令的方式实现。示例性计算设备还可以包括存储在ROM230、RAM240和/或由处理器220执行的其他类型的非暂时性存储介质中的操作系统。程序指令可以与用于提供线上线下服务的操作系统兼容。计算设备200还包括I/O组件260,支持计算机和其他组件之间的输入/输出。计算设备200也可以通过网络通信接收编程和数据。

[0067] 仅用于说明,图2中仅示出了一个处理器。还考虑了多个处理器;因此,由本申请中描述的一个处理器执行的操作和/或方法步骤也可以由多个处理器联合或单独执行。例如,在本申请中,如果计算设备200的处理器执行步骤A和步骤B,应当被理解为是步骤A和步骤B由计算设备200的两个不同的处理器共同或分别执行(例如,第一处理器执行步骤A,第二处理器执行步骤B,或者第一处理器和第二处理器共同执行步骤A和B)。

[0068] 图3是根据本申请的一些实施例所示的示例性移动设备300的示例性硬件和/或软件组件的示意图,在该示例性移动设备300上可以实现用户终端130。

[0069] 如图3所示,移动设备300可以包括通信平台310、显示器320、图形处理单元(GPU) 330、中央处理单元(CPU) 340、I/O350、内存360和存储器390。CPU可以包括接口电路和类似于处理器220的处理电路。在一些实施例中,任何其他合适的组件,包括但不限于系统总线或控制器(未示出),也可以包括在移动设备300内。在一些实施例中,移动操作系统370(例如,iOS™、Android™、Windows Phone™等)和一个或以上应用程序380可以从存储器390加载到内存360中,以便由CPU 340执行。应用程序380可以包括浏览器或任何其他合适的移动应用程序,用于接收和呈现与服务的语音请求有关的信息。用户与信息流的交互可以通过I/O设备350实现,并通过网络120提供给处理引擎112和/或系统100的其他组件。

[0070] 为了实现本申请中描述的各种模块、单元及其功能,计算机硬件平台可以用作一个或多个本文所述元件的硬件平台(例如,线上线下服务AI系统100,和/或线上线下服务AI系统100的其他组件,参考图1-10)。这种计算机的硬件元件、操作系统和编程语言本质上是传统的,并且假设本领域普通技术人员对其进行了充分的熟悉以使这些技术适应于如本文所述的语音请求提供服务响应。一台包含用户界面元素的计算机能够被用作个人计算机(personal computer(PC))或其他类型的工作站或终端设备,被适当程序化后也可以作为服务器使用。可知,本领域技术人员应熟悉该计算机设备的结构、程序设计和一般操作,因此,图对其应是不解自明的。

[0071] 本领域普通技术人员将理解,当线上线下服务AI系统100的元件执行时,该元件可以通过电信号和/或电磁信号执行。例如,当服务器110处理任务时,例如通过一段语音获取服务请求,服务器110可以在其处理器中操作逻辑电路以处理这样的任务。当服务器110接收到语音请求时,服务器110的处理器可以编码语音请求生成电信号。然后,服务器110的处理器可以将电信号发送到与服务器110相关联的目标系统的至少一个信息交换端口。服务

器110通过有线网络与目标系统通信,所述至少一个信息交换端口可以物理连接到电缆,所述电缆还可以将电信号发送到用户终端130的输入端口(例如,信息交换端口)。如果服务器110通过无线网络与目标系统通信,则目标系统的至少一个信息交换端口可以是一个或以上天线,其可以将电信号转换为电磁信号。在诸如用户终端130和/或服务器110的电子设备内,当处理器处理指令、发出指令和/或执行动作时,指令和/或动作通过电信号进行。例如,当处理器从存储介质(例如,存储器140)检索或保存数据时,它可以将电信号发送到存储介质的读/写设备,其可以在存储介质中读取或写入结构化数据。该结构化数据可以用电信号的形式通过电子设备的总线传输至处理器。这里,电信号可以是一个电信号、一系列电信号和/或至少两个分立的电信号。

[0072] 图4是根据本申请的一些实施例所示的示例性处理引擎112的框图。

[0073] 如图4所示,处理引擎112可包括请求获取模块410、模型更新模块420、模型获取模块430、目的地确定模块440和模型训练模块450。

[0074] 请求获取模块410可以被配置为从用户终端130获取语音请求。在一些实施例中,语音请求可以是由一条语音形式的服务请求,其由用户终端130的用户发起。在一些实施例中,语音请求可以是运输服务请求、出租车服务请求、导航服务请求等,或其任何组合。在一些实施例中,语音请求可以是包括目的地、起点、开始时间、服务类型等或其任何组合的一段语音。服务类型可以是运输服务或出租车服务中的车辆类型、导航服务中的旅行模式等,或其任何组合。

[0075] 模型更新模块420可以被配置为更新用于确定语音请求的文字的目的地模型。在一些实施例中,响应于请求获取模块410接收的语音请求,模型更新模块420可以更新定制识别模型。在一些实施例中,可以使用与用户终端相关联的至少两个兴趣点(POI)的数据来训练定制识别模型。训练的定制识别模型可以存储在AI系统100的存储介质(例如,存储器140、ROM230或RAM240)中。当请求获取模块410从用户终端130接收语音请求时,模型更新模块420可以访问存储介质以获取训练的定制识别模型并更新训练的定制识别模型。在一些实施例中,模型更新模块420可以通过在当前语音请求之前输入至少一个历史服务的至少两个POI来更新训练的定制识别模型。例如,模型更新模块420可以通过将用户和/或用户终端的至少一个历史服务的至少一个起点和至少一个目的地输入到训练的定制识别模型中来更新训练的定制识别模型。在一些实施例中,训练的定制识别模型的更新可以是调整或修改训练的定制识别模型中的参数的过程。例如,训练的定制识别模型的更新可以导致语音请求的识别结果比训练的定制识别模型更准确。在一些实施例中,更新的定制识别模型可以与训练的定制识别模型相同。例如,用于更新训练的定制识别模型的至少一个起点和至少一个目的地的数据可以与用于训练定制识别模型的至少两个POI的数据相同。

[0076] 在一些实施例中,模型更新模块420可以实时在线更新训练的定制识别模型。例如,每当请求获取模块410接收到新的语音请求时,模型更新模块420可以在当前语音请求之前输入至少一个历史服务的至少两个POI以更新训练的定制识别模型。在一些实施例中,模型更新模块420可以实时更新训练的定制识别模型,并且模型更新模块420还可以周期性地更新训练的定制识别模型。例如,模型更新模块420可以在预定时间段(例如,每几个小时、每天、每几天、每个月等)的时间间隔更新训练的定制识别模型。又例如,模型更新模块420可以在用户终端的用户完成服务之后更新训练的定制识别模型。

[0077] 模型获取模块430可以被配置为获取模型。在一些实施例中,模型获取模块430可以从AI系统100的存储介质(例如,存储器140、ROM230、RAM240或存储模块)获取通用识别模型和/或定制识别模型。在一些实施例中,通用识别模型和/或定制识别模型可以由模型训练模块450训练。

[0078] 目的地确定模块440可以被配置为确定与从请求获取模块410获取的语音请求相关联的文字的目的地。

[0079] 在一些实施例中,目的地确定模块440至少基于语音请求、定制识别模型和通用识别模型来确定文字的目的地。目的地确定模块440可以基于语音请求和定制识别模型确定至少一个定制结果,并且基于语音请求和通用识别模型确定至少一个通用结果。目的地确定模块440可以从至少一个定制结果和至少一个通用结果确定作为文字的目的的候选文字序列。

[0080] 在一些实施例中,目的地确定模块440可以至少基于语音请求、定制语音模型和语言识别模型来确定文字的目的地。例如,目的地确定模块440可以基于声学模型和语音请求确定语音请求的至少两个声学结果,并至少基于至少两个声学结果、定制语音模型和通用语音模型确定文字的目的地。

[0081] 模型训练模块450可以被配置为训练模型。在一些实施例中,模型训练模块450可以训练定制识别模型。例如,模型训练模块450可以获取与用户终端相关联的至少两个兴趣点(POI)的数据,并且将至少两个POI的数据输入到用于将语音识别为文本的算法和/或方法中(例如,初始语音识别模型、初始语音模型)以获取定制识别模型。在一些实施例中,模型训练模块450可以训练通用识别模型。例如,模型训练模块450可以从公众获取至少两个位置的数据,并且将至少两个来自公众位置的数据输入到用于将语音识别为文本的算法和/或方法(例如,初始语音识别模型、初始语音模型)以获取通用识别模型。

[0082] 处理引擎112中的模块可以通过有线连接或无线连接彼此连接或通信。有线连接可以是金属线缆、光缆、混合线缆等,或其任何组合。无线连接可以是局域网络(LAN)、广域网络(WAN)、蓝牙、紫蜂网络、近场通讯(NFC)等,或其任意组合。两个或以上模块可以被组合为单个模块,且所述模块中的任一个可以被分成两个或以上单元。例如,模型更新模块420和模型训练模块450可以组合为单个模块,其可以训练模型并更新模型。又例如,目的地确定模块440可以分为定制结果判断单元、通用结果判断单元和文字序列判断单元,其可以分别确定定制结果、通用结果和作为文字的目的的候选文字序列。作为又一示例,处理引擎112可包括用于存储模型和/或文字的目的地数据和/或信息的存储模块(未示出)。

[0083] 图5A是根据本申请的一些实施例所示的确定文字的目的的示例性过程的流程图。过程500可以由线上线下服务AI系统100执行。例如,过程500可以作为存储在存储ROM230或RAM240中的一组指令(例如,应用程序)来实现。处理器220可以执行该组指令,并且当执行指令时,可以将其配置为执行过程500。以下所示过程的操作仅出于说明的目的。在一些实施例中,过程500可以利用未描述的一个或以上附加操作,和/或没有所讨论的一个或以上操作来完成。另外,如图5A所示和下面描述的过程操作的顺序不是限制性的。

[0084] 在过程510中,处理引擎112(例如,处理器220、请求获取模块410)可以从用户终端130接收语音请求。在一些实施例中,语音请求可以是一段语音形式的服务请求,其由用户终端130的用户发起。在一些实施例中,语音请求可以是运输服务请求、出租车服务请求、导

航服务请求等,或其任何组合。在一些实施例中,语音请求可以是与服务相关联的一段语音,包括但不限于目的地、起点、开始时间、服务类型等,或其任何组合。服务类型可以是运输服务或出租车服务中的车辆类型、导航服务中的旅行模式等,或其任何组合。

[0085] 在一些实施例中,用户终端130的用户可以通过无线通信将语音请求发送和/或传输到目标系统的信息交换端口。目标系统可以包括用户终端130、网络120,以及网络120和服务器110之间的信息交换端口。信息交换端口可以是网络120和服务器110之间的接口。例如,信息交换端口可以是连接到服务器110和/或与服务器110通信的一个或以上网络I/O端口(例如,天线)。目标系统的信息交换端口可以将语音请求发送到处理引擎112。在一些实施例中,用户可以是乘客、司机、旅行者等,或其任何组合。

[0086] 在过程520中,响应于用户终端130接收的语音请求,处理引擎112(例如,处理器220、模型更新模块420)可以更新定制识别模型。

[0087] 在一些实施例中,可以使用与用户终端相关联的至少两个兴趣点(POI)的数据来训练定制识别模型。例如,处理引擎112可以将至少两个POI的数据输入到用于将语音识别为文本的算法和/或方法中(例如,初始语音识别模型、初始语音模型)以获取定制识别模型。

[0088] 在一些实施例中,与用户终端相关联的POI可以包括用户终端的用户曾经使用和/或定位的至少两个位置。例如,POI可以包括从用户接收的至少两个历史服务获取的至少两个位置,包括观光历史、餐饮和娱乐历史、运输服务历史,例如出租车服务、导航服务、货物递送服务、观光历史等,或其任何组合。POI可以包括至少两个历史服务的至少两个历史出发点、至少两个历史服务的至少两个历史目的地、用户在至少两个历史服务期间或其任何组合中所定位的至少两个历史位置。

[0089] 在一些实施例中,与用户终端相关联的POI可以包括与用户和/或用户终端的特定主题相关联的至少两个位置。例如,POI可以包括用户和/或用户终端所在区域中的至少两个位置,至少两个类似用户的至少两个历史位置(例如,历史出发点、历史目的地等),或其任何组合。该区域可以是城市区域、地区区域、预定区域等,或其任何组合。类似用户可以是具有相同或相似的旅行或导航习惯、做相同或类似的工作、年龄相近等,或其任何组合的至少两个用户。在一些实施例中,处理引擎112对AI系统100的至少两个用户进行分类,以获取与用户终端相关联的类似用户。

[0090] 在一些实施例中,用于训练定制识别模型的至少两个POI的数据可以是至少两个POI的文字信息、至少两个POI的发音,或至少两个POI的任何语言特征,或其任何组合。

[0091] 在一些实施例中,定制识别模型可以是识别定制识别模型的输入,并将输入转换为文本、代码或可由人类用户或电子设备识别的其他书面形式的AI算法和/或方法。定制识别模型可以特定于特定个体,例如用户终端的用户。

[0092] 在一些实施例中,定制识别模型的输入可以是语音请求,定制识别模型可以是定制AI算法和/或方法,其可以将语音请求中的语音识别为文本、代码或人类用户或电子设备可识别的其他书写形式(例如,文字的目的地、文字来源、文字服务类型等)。例如,处理引擎112可以将用于请求出租车呼叫服务的语音请求输入到定制识别模型中。定制识别模型的输出可以是语音请求识别的文本结果。处理引擎112可以提取包括在文本结果中的目标POI。在一些实施例中,定制识别模型可包括声学模型、语音模型、发音词典、解码器等,或其

任何组合。在一些实施例中，定制识别模型可以是由至少两个POI的文字信息和至少两个POI的相应发音训练的集成AI模型。当处理引擎112将语音请求输入到定制识别模型时，定制识别模型可以输出从语音请求识别的文本结果。处理引擎112可以提取包括在文本结果中的目标POI。使用定制识别模型将语音请求识别到文字的目的地过程的示意图可以在图6A及其描述中找到。

[0093] 在一些实施例中，定制识别模型的输入可以从语音请求获取的信息。例如，处理引擎112可以将语音请求输入到声学模型中以获取至少两个声学结果（例如，语音元素、语言单元、音素、音节等）。定制识别模型可以是定制语音模型，其可以将语音请求的声学结果识别为文本（例如，文字的目的地、文字来源、文字服务类型等）。例如，处理引擎112可以使用发音词典将来自声学模型的语音请求输出的至少两个音素映射到单词。处理引擎112可以将至少两个音素的映射单词输入到定制识别模型中。定制识别结果可以包括至少一个定制文字序列和相应的序列概率，所述序列概率表示定制文字序列与语音请求相关联的概率。例如，序列概率可以是用户可以在语音请求中表达对应的定制文字序列的可能性。在一些实施例中，处理引擎112可以基于声学模型和定制识别模型确定序列概率。例如，序列概率可以是对应于从声学模型输出的至少两个音素的声学概率乘积与对应于从定制识别模型输出的至少两个单词的文字概率的乘积之和。处理引擎112可以通过从至少一个定制文字序列中选择具有最大序列概率的定制文字序列来以文本的形式确定用户的起点和/或目的地。在一些实施例中，定制语音模型可以包括或可以由unigram模型、N-Gram模型、指数语音模型、神经语音模型等，或其任何组合训练的模型。使用定制语音模型将语音请求识别到文字的目的地过程的示意图可以在图6B及其描述中找到。

[0094] 在一些实施例中，训练的定制识别模型可以存储在AI系统100的存储介质（例如，存储器140、ROM230或RAM240）中。当处理引擎112从用户终端130接收语音请求时，处理引擎112可以访问存储介质以获取训练的定制识别模型并更新训练的定制识别模型。在一些实施例中，处理引擎112可以通过在当前语音请求之前输入至少一个历史服务的至少两个POI来更新训练的定制识别模型。例如，处理引擎112可以通过将用户和/或用户终端的至少一个历史服务的至少一个起点和至少一个目的地输入到训练的定制识别模型中来更新训练的定制识别模型。在一些实施例中，训练的定制识别模型的更新可以是调整或修改训练的定制识别模型中的参数的过程。训练的定制识别模型的更新可以导致语音请求的识别结果比训练的定制识别模型更准确。在一些实施例中，更新的定制识别模型可以与训练的定制识别模型相同。例如，用于更新训练的定制识别模型的至少一个起点和至少一个目的地的数据可以与用于训练定制识别模型的至少两个POI的数据相同。

[0095] 在一些实施例中，处理引擎112可以实时在线更新训练的定制识别模型。例如，每当处理引擎112接收到新的语音请求时，处理引擎112可以在当前语音请求之前将至少一个历史服务的至少两个POI输入到训练的定制识别中以更新训练的定制识别模型。在一些实施例中，处理引擎112可以实时更新训练的定制识别模型，并且还可以周期性地更新训练的定制识别模型。例如，处理引擎112可以在预定时间段（例如，每几个小时、每天、每几天、每个月等）的时间间隔更新训练的定制识别模型。又例如，处理引擎112可以在用户终端的用户完成服务之后更新训练的定制识别模型。

[0096] 在一些实施例中，响应于语音请求，处理引擎112可在线训练初始识别模型以获取

定制识别模型。例如,每当处理引擎112接收到新的语音请求时,处理引擎112可以在当前语音请求之前将用户和/或用户终端的至少两个历史服务的所有POI输入到初始识别模型中。初始识别模型可以是用于将语音识别为文本的算法和/或方法。

[0097] 在过程530中,处理引擎112(例如,处理器220、模型获取模型430)可以获取通用识别模型。在一些实施例中,可以使用从公众收集的数据来训练通用识别模型。例如,处理引擎112可以将从公众收集的数据输入到用于将语音识别为文本的算法和/或方法中(例如,初始识别模型、初始语音模型)以获取通用识别模型。通用识别模型可以与定制模型具有相同类型的模型。例如,可以使用从不同源收集的数据从相同的初始模型训练通用识别模型和定制识别模型。或者,通用识别模型和定制识别模型可以是不同类型的模型。

[0098] 在一些实施例中,来自公众的数据可以包括来自公众的至少两个位置的文字信息、来自公众的至少两个位置的发音、或来自公众的至少两个位置的任何语言特征,或其任何组合。在一些实施例中,来自公众的至少两个位置可以由网络爬虫在整个因特网、网站、服务服务器等或其任何组合中爬行。

[0099] 在一些实施例中,通用识别模型可以是识别通用识别模型的输入并将输入转换为文本、代码或可由人类用户或电子设备识别其他书面形式的AI算法和/或方法。通用识别模型对于公众的所有个人都是通用。

[0100] 在一些实施例中,通用识别模型的输入可以是语音请求,并且通用识别模型可以是AI算法和/或方法,其可以将语音请求中的语音识别为文本、代码或人类用户或电子设备可识别的其他书写形式(例如,文字的目的地、文字来源、文字服务类型等)。例如,处理引擎112可以将用于请求出租车呼叫服务的语音请求输入到通用识别模型中。通用识别模型的输出可以是语音请求识别的文本结果。处理引擎112可以提取包括在文本结果中的目标POI。在一些实施例中,通用识别模型可包括声学模型、语音模型、发音字典、解码器等,或其任何组合。在一些实施例中,通用识别模型可以是集成的AI模型,其由来自公众的至少两个位置的文本数据和至少两个位置的相应发音进行训练。当处理引擎112将语音请求输入到通用识别模型时,通用识别模型可以输出从语音请求识别的文本结果。处理引擎112可以提取包括在文本结果中的目标POI。使用通用识别模型将语音请求识别到文字的目的地过程的示意图可以在图6A及其描述中找到。

[0101] 在一些实施例中,通用识别模型的输入可以是语音请求获取的信息。例如,处理引擎112可以将语音请求输入到声学模型中以获取至少两个声学结果(例如,语音元素、语言单元、音素、音节等)。通用识别模型可以是通用语音模型,其可以将语音请求的声学结果识别为文本(例如,文字的目的地、文字来源、文字服务类型等)。例如,处理引擎112可以使用发音字典将来自声学模型输出的至少两个音素映射到单词,并且将至少两个音素的映射结果输入到通用识别模型中。公认的结果可以包括至少一个通用文字序列和相应的序列概率,所述序列概率表示通用文字序列与语音请求相关联的概率。例如,序列概率可以是用户可以在语音请求中表达相应的通用文字序列的可能性。在一些实施例中,处理引擎112可以基于声学模型和通用识别模型确定序列概率。例如,序列概率可以是对应于从声学模型输出的至少两个音素的声学概率乘积与对应于从通用识别模型输出的至少两个单词的文字概率的乘积之和。处理引擎112可以通过从至少一个通用文字序列中选择具有最大序列概率的通用文字序列来以文本的形式确定用户的起点和/或目的地。在一些实施例中,通

用语音模型可以由unigram模型、N-Gram模型、指数语音模型、神经语音模型等或其任何组合训练的模型。使用通用语音模型将语音请求识别到文字的目的地过程的示意图可以在图6B及其描述中找到。

[0102] 对于不同的语音请求,通用识别模型和定制识别模型可以显示不同的识别准确度。例如,如果语音请求包括用户曾经使用和/或定位的位置,或者语音请求的用户具有强烈的重音,定制识别模型可以识别比通用识别模型的识别结果更准确的文字结果。如果语音请求包括用户从未使用和/或定位的位置,则由于缺少训练数据,来自定制识别模型的识别结果可能具有比来自通用识别模型的识别结果更低的识别精度。

[0103] 在过程540中,处理引擎112(例如,处理器220、目的地确定模块440)可以至少基于语音请求、定制识别模型和通用识别模型确定与语音请求相关联的文字的目的地。

[0104] 文字的目的地可以通过文本、代码或智能设备可识别的任何书面形式表达的地方和/或可转换为人类用户可识别的书面形式。仅用于说明目的,当前本申请使用文本作为文字的目的地。在一些实施例中,文字的目的地可以从语音请求识别的位置的任何文本信息。例如,文字的目的地可以是服务的起点的文本信息、服务的目的地的文本信息、服务期间的通过位置的文本信息等,或其任何组合。

[0105] 在一些实施例中,处理引擎112可以基于语音请求和定制识别模型确定至少一个定制结果,并且基于语音请求和通用识别模型确定至少一个通用结果。处理引擎112可以比较至少一个定制结果和至少一个通用结果,并从至少一个定制结果和至少一个通用结果中选择最佳结果作为文字的目的地。最佳结果可以使具有文字的目的地的位置,其具有用户可以在语音请求中表达的最大可能性。

[0106] 在一些实施例中,所述至少一个定制结果可包括至少一个定制文字序列及其相应的序列概率。至少一个通用结果可以包括至少一个通用文字序列和其相应的序列概率。文字序列(例如,定制文字序列或通用文字序列)可以是预测的文字序列,所述文字序列是定制识别模型或通用识别模型预测的用户可以在语音请求中表达的文字序列。对应的序列概率可以是用户可以在语音请求中表达对应的定制文字序列的可能性。序列概率越高,相应的定制文字序列越可能是相应模型从语音请求识别出的用户表达。

[0107] 在一些实施例中,处理引擎112可以比较至少一个定制文字序列的序列概率和至少一个通用文字序列的序列概率,并从所述至少一个定制文字序列和所述至少一个通用文字序列中选择具有最大序列概率的候选文字序列。候选文字序列可以被指定为文字的目的地。

[0108] 例如,处理引擎112从用户接收包括“鼎好大厦”的语音信息的语音请求。用户曾使用“鼎好大厦”作为出租车服务的历史目的地。定制识别模型使用“鼎好大厦”的数据进行训练。当处理引擎112分别将语音请求(或从语音请求获取的数据)输入到定制识别模型和通用识别模型时,定制识别模型输出具有序列概率为99%的文字“鼎好大厦”;通用识别模型以序列概率为90%输出文字“帝豪大厦”,以及序列概率为80%的文字“鼎好大厦”。处理引擎112可以指定具有从定制识别模型输出的99%的序列概率作为文字的目的地文字“鼎好大厦”。

[0109] 又例如,处理引擎112从用户接收包括“鼎好大厦”的语音信息的语音请求。用户从未使用“鼎好大厦”作为历史出发点或出租车服务中的目的地。当处理引擎112分别将语音

请求(或从语音请求获取的数据)输入到定制识别模型和通用识别模型时,定制识别模型输出文字“鼎好大厦”,其序列概率为69%;通用识别模型以序列概率为95%输出文字“帝豪大厦”,以及序列概率为80%的文字“鼎好大厦”。处理引擎112可以指定文字“帝豪大厦”,其具有从通用识别模型输出的90%的序列概率作为文字的目的地。

[0110] 在一些实施例中,在确定文字的目的地之后,处理引擎112(例如,处理器220)可以生成包括文字的目的地和触发代码的电子信号。在一些实施例中,触发代码可以由安装在用户终端130中的应用程序(例如,运输服务应用程序、出租车服务应用程序、导航服务应用程序等)正式识别。例如,触发代码可以是指令、代码、标记、符号等,或其任何组合,其可以激励用户终端130或让用户终端130执行任何程序。触发代码可以被配置为使应用程序在用户终端130的界面上生成文字的目的地。图5B是根据本申请的一些实施例所示的用户终端130的示例性用户界面,用于呈现文字的目的地。文字“兴化巷”和“鼎好大厦”被认为是文本的起点和目的地。在一些实施例中,处理引擎112可以将电子信号发送到目标系统的信息交换端口,以使信息交换端口将电子信号发送到用户终端130。响应于接收电子信号,用户终端130可以在其界面上呈现文字的目的地,以让用户确认文字的目的地是否是用户可以表达的。用户可以在框中单击“是”和“否”的两个图标中的任何一个,以确认文字的目的地是否是用户表达的。

[0111] 图6A是根据本申请的一些实施例所示的将语音请求识别到文字的目的地示例性过程的示意图。

[0112] 如图6A所示,语音请求可以分别输入到定制识别模型和通用识别模型中。定制识别模型的输出可以包括至少两个定制文字序列和至少两个定制文字序列的相应序列概率,即,从定制识别模型输出的每个序列概率对应于定制文字序列。通用识别模型的输出可以包括至少两个通用文字序列和至少两个通用文字序列的相应序列概率,即,从通用识别模型输出的每个序列概率对应于通用文字序列。系统可以比较来自所述2模型的序列概率以获取来自至少两个定制文字序列的文字序列和至少两个通用文字序列中的具有最大序列概率的文字序列作为文字的目的地。确定文字的目的地细节可以在本申请的其他地方找到(例如,图7及其描述)。

[0113] 图6B是根据本申请的一些实施例所示的将语音请求识别到文字的目的地示例性过程的示意图。

[0114] 如图6B所示,语音请求可以输入到特征提取器中。特征提取器可以从语音请求中提取至少两个特征。提取的特征可以输入到解码器中。解码器可以包括声学模型、字典、定制语音模型和通用语音模型。提取的特征可以输入到声学模型中,并且声学模型可以输出至少两个音素以及至少两个音素的相应声学概率。可以将至少两个音素输入到词典中以查找对应于词典中的至少两个音素的至少两个单词。可以将至少两个单词输入到定制语音模型和通用语音模型中。定制结果可包括至少两个定制文字序列和至少两个定制文字序列的相应序列概率,并且通用结果可以包括至少两个通用文字序列和至少两个通用文字序列的相应序列概率。解码器可以比较至少两个定制文字序列的序列概率和至少两个通用文字序列的序列概率,并输出来自至少两个定制文字序列和至少两个通用文字序列中具有最大序列概率的文字序列作为文字的目的地。将语音请求识别到文字的目的地细节可以在本申请的其他地方找到(例如,图7-10及其描述)。

[0115] 图7是根据本申请的一些实施例所示的确定文字的目的地的示例性过程的流程图。过程700可以由线上线下服务AI系统100执行。例如,过程700可以实现为存储在存储ROM230或RAM240中的一组指令(例如,应用程序)。处理器220可以执行该组指令,并且当执行该指令时,可以将其配置为执行该过程700。以下所示过程的操作仅出于说明的目的。在一些实施例中,过程700可以利用未描述的一个或以上附加操作和/或没有所讨论的一个或以上个操作来完成。另外,如图7所示和下面描述的过程操作的顺序不是限制性的。在一些实施例中,可以基于过程700来执行过程500中的步骤540。

[0116] 在过程710中,处理引擎112(例如,处理器220、目的地确定模块440)可以基于语音请求和定制识别模型确定至少一个定制结果。所述至少一个定制结果中的每一个可包括定制文字序列和序列概率,所述序列概率表示语音请求与定制文字序列相关联的概率。

[0117] 在一些实施例中,处理引擎112可以将语音请求输入到定制识别模型中,如图6A所示。定制识别模型可以分析语音请求,并输出至少一个定制文字序列和相应的序列概率。定制文字序列可以是用户可以基于定制识别模型在语音请求中表达的预测文字序列。对应的序列概率可以预测用户可以在语音请求中表达对应的定制文字序列的可能性。序列概率越高,用户可以在语音请求中表达的相应定制文字序列的可能性越大。

[0118] 在一些实施例中,定制识别模型可以是定制语音模型。处理引擎112可以处理语音请求,并将语音请求的处理结果输入到定制语音模型中,如图6B所示。例如,处理引擎112可以使用特征提取器从语音请求中提取至少两个声学特征。至少两个声学特征可包括光谱特征、时域特征、梅尔频率倒谱系数(MFCC)特征、对数梅尔频谱系数(MFSC)特征等,或其任何组合。处理引擎112可以将至少两个声学特征输入到声学模型中以获取至少两个候选语音元素和相应的声学概率,所述概率表示语音请求与候选语音元素相关联的概率。至少两个候选语音元素可包括音素、音节或任何语言单元,或其任何组合。至少两个候选语音元素可以用至少两个候选语音元素映射字典中的单词,并将映射单词输入到定制语音模型中。定制语音模型可以分析映射单词以获取至少一个定制文字序列和相应的序列概率。

[0119] 在过程720中,处理引擎112(例如,处理器220、目的地确定模块440)可以基于语音请求和通用识别模型确定至少一个通用结果。至少一个通用结果中的每一个可以包括通用文字序列和序列概率,所述序列概率表示语音请求与通用文字序列相关联的概率。

[0120] 在一些实施例中,处理引擎112可以将语音请求输入到通用识别模型中,如图6A所示。通用识别模型可以分析语音请求,并输出至少一个通用文字序列和相应的序列概率。通用文字序列可以是用户可以基于通用识别模型在语音请求中表达的预测文字序列。相应的序列概率可以预测用户可以在语音请求中表达相应的通用文字序列的可能性。序列概率越高,用户可以在语音请求中表达的相应通用文字序列的可能性越大。

[0121] 在一些实施例中,通用识别模型可以是通用语音模型。处理引擎112可以处理语音请求,并将语音请求的处理结果输入到通用语音模型中,如图6B所示。处理语音请求以及获取至少一个通用文字序列和对应的序列概率的方法和/或过程可以与过程710中描述的方法和/或过程类似。

[0122] 在过程730中,处理引擎112(例如,处理器220,目的地确定模块440)可以从所述至少一个定制文字序列和所述至少一个通用文字序列确定一个候选文字序列作为文字的目的地。候选文字序列可以是至少一个定制文字序列和至少一个通用文字序列中的最大序列

概率。

[0123] 在一些实施例中,处理引擎112可以比较至少一个定制文字序列的序列概率和至少一个通用文字序列的序列概率,并从所述至少一个定制文字序列和所述至少一个通用文字序列中选择具有最大序列概率的候选文字序列。候选文字序列可以被指定为文字的目的地。

[0124] 例如,处理引擎112从用户接收包括“鼎好大厦”的语音信息的语音请求。用户曾使用“鼎好大厦”作为出租车服务的历史目的地。定制识别模型使用“鼎好大厦”的数据进行训练。当处理引擎112分别将语音请求(或从语音请求获取的数据)输入到定制识别模型和通用识别模型时,定制识别模型输出具有序列概率为99%的文字“鼎好大厦”;通用识别模型以序列概率为90%输出文字“帝豪大厦”,以及序列概率为80%的文字“鼎好大厦”。处理引擎112可以指定具有从定制识别模型输出的99%的序列概率作为文字的目的地文字“鼎好大厦”。

[0125] 又例如,处理引擎112从用户接收包括“鼎好大厦”的语音信息的语音请求。用户从未使用“鼎好大厦”作为历史出发点或出租车服务中的目的地。当处理引擎112分别将语音请求(或从语音请求获取的数据)输入到定制识别模型和通用识别模型时,定制识别模型输出文字“鼎好大厦”,其序列概率为69%;通用识别模型以序列概率为95%输出文字“帝豪大厦”,以及序列概率为80%的文字“鼎好大厦”。处理引擎112可以指定文字“帝豪大厦”,其具有从通用识别模型输出的90%的序列概率作为文字的目的地。

[0126] 在一些实施例中,处理引擎112可以为定制识别模型的定制结果和通用识别模型的通用结果分配不同的权重,并通过计算至少一个定制文字序列和至少一个通用文字序列的加权概率来确定文字的目的地。

[0127] 图8是根据本申请的一些实施例所示的确定文字的目的地示例性过程的流程图。过程800可以由线上线下服务AI系统100执行。例如,过程800可以实现为存储在存储ROM230或RAM240中的一组指令(例如,应用程序)。处理器220可以执行该组指令,并且当执行指令时,可以将其配置为执行过程800。以下所示过程的操作仅出于说明的目的。在一些实施例中,过程800可以利用未描述的一个或以上附加操作和/或没有所讨论的一个或以上个操作来完成。另外,如图8所示和下面描述的过程操作的顺序不是限制性的。

[0128] 在过程810中,处理引擎112(例如,处理器220,文字的目的地确定模块440)可以基于声学模型和语音请求确定语音请求的至少两个声学结果。

[0129] 在一些实施例中,处理引擎112可以处理语音请求(例如,提取声学特征),并将语音请求的处理结果输入到声学模型中。声学模型可以是预测构成语音请求的语音的可能(或候选)语音元素(例如,音素、音节、语言单元)的方法和/或算法。声学模型可以输出至少两个声学结果。声学结果可以是声学模型预测的预测结果。声学结果可包括至少两个候选语音元素,例如音素、音节或任何语言单元,或其任何组合。确定至少两个声学结果的细节可以在本申请的其他地方找到(例如,图9及其描述)。

[0130] 在过程820中,处理引擎112(例如,处理器220、文字的目的地确定模块440)可以确定文字的目的地至少有一个至少两个声学结果、定制语音模型和通用语音模型。

[0131] 在一些实施例中,可以在本申请的过程500的过程520中获取定制语音模型。通用语音模型可以在本申请的过程500的过程530中获取。在一些实施例中,处理引擎112可以至

少处理至少两个声学结果以通过字典获取至少两个映射单词。然后,处理引擎112可以将至少两个映射单词分别输入到定制语音模型和通用语音模型中。确定文字的目的地的方法和/或过程可以与本申请中的过程700类似。

[0132] 图9是根据本申请的一些实施例所示的确定至少两个声学结果的示例性过程的流程图。过程900可以由线上线下服务AI系统100执行。例如,过程900可以实现为存储在存储ROM230或RAM240中的一组指令(例如,应用程序)。处理器220可以执行该组指令,并且当执行该指令时,可以将其配置为执行该过程900。以下所示过程的操作仅出于说明的目的。在一些实施例中,过程900可以利用未描述的一个或以上附加操作和/或没有所讨论的一个或以上个操作来完成。另外,如图9所示和下面描述的过程操作的顺序不是限制性的。在一些实施例中,可以基于过程900来执行过程800中的步骤810。

[0133] 在过程910中,处理引擎112(例如,处理器220、目的地确定模块440)可以获取语音请求的至少两个帧。在一些实施例中,处理引擎112可以按时间顺序将语音请求划分为至少两个帧。至少两个帧中的每一个可以包括预定时间段的数据。例如,每个帧可以包括25毫秒、100微秒、25毫秒、50毫秒等的数据,或其任何组合。

[0134] 在过程920中,对于至少两个获取的帧的每个帧,处理引擎112(例如,处理器220、目的地确定模块440)可以识别至少一个声学特征。在一些实施例中,处理引擎112可以通过特征提取器从每个帧中提取至少一个声学特征。在一些实施例中,声学特征可包括光谱特征、时域特征、梅尔频率倒谱系数(MFCC)特征、对数梅尔频谱系数(MFSC)特征等,或其任何组合。

[0135] 在过程930中,对于至少两个获取的帧的每个帧,处理引擎112(例如,处理器220、目的地确定模块440)可以基于声学模型确定声学结果,其中,所述至少一个声学特征是声学模型的输入。

[0136] 声学结果可以是声学模型预测的预测结果。例如,声学结果可以表示音频信号和语言单元(例如,语音元素、音素、音节)之间的关系。在一些实施例中,对于每个帧,处理引擎112可以将帧的至少一个声学特征输入到声学模型中。声学结果可以是声学模型的输出。在一些实施例中,声学结果可包括至少一个候选语音元素和相应的声学概率,所述概率表示语音请求与候选语音元素相关联的概率。例如,声学概率可以预测语音请求的每个帧可以发音相应的候选语音元素的可能性。声学概率越大,语音请求的相应帧发出相应的候选语音元素的可能性越大。在一些实施例中,至少两个候选语音元素可包括构成语音请求的帧的音素、音节或任何语言单元,或其任何组合。

[0137] 图10是根据本申请的一些实施例所示的确定文字的目的的示例性过程的示意图。

[0138] 在一些实施例中,对于语音请求的每个帧,处理引擎112(例如,处理器220、目的地确定模块440)可以基于该帧的至少一个声学特征和声学模型获取至少两个候选语音元素。例如,处理引擎112可以将至少一个声学特征输入到声学模型中。声学模型可输出至少两个候选语音元素和相应的声学概率,所述概率表示语音请求与候选语音元素相关联的概率。声学概率可以预测语音请求的每个帧可以发音相应的候选语音元素的可能性。声学概率越大,语音请求的相应帧发出相应的候选语音元素的可能性越大。在一些实施例中,至少两个候选语音元素可包括构成语音请求的帧的音素,音节或任何语言单元,或其任何组合。

[0139] 如图10所示,处理引擎112可以通过至少两个路径从“开始”节点到“结束”节点进行第一帧。例如,在一个路径中,声学模型输出候选语音元素,例如 $A_1$ 、 $A_2$ 、 $A_3$ 和 $A_{11}$ ,以及它们相应的声学概率。然后,处理引擎112可以查找字典得到 $A_1$ 、 $A_2$ 和 $A_3$ 对应于单词 $W_1$ ,并且定制语音模型(或通用语音模型)可以输出 $W_1$ 的文字概率。文字概率可以是用户可以在语音请求的每个帧中表达相应单词的可能性。定制语音模型(或通用语音模型)输出的文字概率越大,相应的单词越可能是用户在帧中表达的。又例如,在另一条路径中,声学模型输出候选语音元素,例如 $A_1$ 、 $A_4$ 、 $A_5$ 和 $A_6$ ,以及它们相应的声学概率。然后,处理引擎112可以查找字典得到 $A_1$ 、 $A_4$ 、 $A_5$ 和 $A_6$ 对应于单词 $W_2$ ,并且定制语音模型(或通用语音模型)可以输出 $W_2$ 的文字概率。省略多个路径。在继续第一帧之后,处理引擎112然后可以使用相同的方法进行下一帧。处理引擎112可以从每个帧中选择一个路径并依次链接语音请求的至少两个帧的至少两个相应路径。表示用户在语音请求中表达相应的文字序列的可能性的序列概率可以是至少两个声学概率的乘积与对应于语音请求的链接路径的至少两个文字概率的乘积之和。例如,处理引擎112可以从每个帧中选择一个路径并添加相应的路径概率以获取序列概率。处理引擎112可以选择具有最大序列概率的文字序列作为对应于语音请求的文字的目的地。

[0140] 图11是根据本申请的一些实施例所示的训练通用识别模型的示例性过程的流程图。过程1100可以由线上线下服务AI系统100执行。例如,过程1100可以实现为存储在存储ROM230或RAM240中的一组指令(例如,应用程序)。处理器220可以执行该组指令,并且当执行该指令时,可以将其配置为执行该过程1100。以下所示过程的操作仅出于说明的目的。在一些实施例中,过程900可以利用未描述的一个或以上附加操作和/或没有所讨论的一个或以上个操作来完成。另外,如图11所示和下面描述的过程操作的顺序不是限制性的。在一些实施例中,可以基于过程1100来执行过程500中的步骤530。

[0141] 在过程1110中,处理引擎112(例如,处理器220、模型训练模块450)可以从公众获取至少两个位置的文字信息。

[0142] 在一些实施例中,至少两个位置的文字信息可以是文本形式的至少两个位置名称。在一些实施例中,来自公众的至少两个位置可以由网络爬虫在整个因特网、网站、服务器等或其任何组合中爬行。例如,来自公众的至少两个位置可以包括特定服务平台的用户已经使用的历史位置,诸如历史目的地、历史出发点、历史服务位置等,或其任何组合。又如例,来自公众的至少两个位置可以是出现在万维网(WWW)中的位置。

[0143] 在过程1120中,处理引擎112(例如,处理器220、模型训练模块450)可以使用至少两个位置的文字信息训练初始语音模型作为初始语音模型的输入以获取通用识别模型。在一些实施例中,通用识别模型可以是通用语音模型。

[0144] 在一些实施例中,初始语音模型可以是用于预测单词出现在文字序列(例如,短语、句子等)中的可能性的算法和/或方法。例如,初始语音模型可以是单字模型、N-Gram模型、指数语音模型、神经语音模型等,或其任何组合。

[0145] 在一些实施例中,处理引擎112可以将公众的至少两个位置输入到初始语音模型中以训练初始语音模型。例如,处理引擎112可以以从出租车服务平台爬行的文本的形式将至少两个历史位置名称输入到神经语音模型中。可以确定神经语音模型中的参数(例如,层、节点、权重等)以获取通用语音模型。

[0146] 在一些实施例中,处理引擎112可以线下训练通用识别模型。例如,处理引擎112可

以每隔预定时间段训练通用识别模型。又例如,处理引擎112可以在爬行至少两个位置的文字信息之后仅训练一次通用识别模型。在一些实施例中,处理引擎112可以将训练好的通用识别模型存储在AI系统100的存储介质(例如,存储器140、ROM230或RAM240)中。

[0147] 上文已对基本概念做了描述,显然,对于阅读此申请后的本领域的普通技术人员来说,上述发明披露仅作为示例,并不构成对本申请的限制。虽然此处并未明确说明,但本领域的普通技术人员可能会对本申请进行各种修改、改进和修正。该类修改、改进和修正在本申请中被建议,所以该类修改、改进、修正仍属于本申请示范实施例的精神和范围。

[0148] 同时,本申请使用了特定词语来描述本申请的实施例。例如“一个实施例”、“一实施例”、和/或“一些实施例”意指与本申请至少一个实施例相关的某一特征、结构或特性。因此,应强调并注意的是,本说明书中在不同位置两次或以上提及的“一实施例”或“一个实施例”或“一替代性实施例”并不一定是指同一实施例。此外,本申请的一个或以上实施例中的某些特征、结构或特点可以进行适当的组合。

[0149] 此外,本领域的普通技术人员可以理解,本申请的各方面可以通过若干具有可专利性的种类或情况进行说明和描述,包括任何新的和有用的过程、机器、产品或物质的组合,或对其任何新的和有用的改进。相应地,本申请的各个方面可以完全由硬件执行、可以完全由软件(包括固件、常驻软件、微码等)执行、也可以由硬件和软件组合执行。以上硬件或软件均可被称为“数据块”、“模块”、“引擎”、“单元”、“组件”或“系统”。此外,本申请的各方面可以采取体现在一个或以上计算机可读介质中的计算机程序产品的形式,其中计算机可读程序代码包含在其中。

[0150] 计算机可读信号介质可能包含一个内含有计算机程序代码的传播数据信号,例如在基带上或作为载波的一部分。此类传播信号可以有多种形式,包括电磁形式、光形式等或任何合适的组合。计算机可读信号介质可以是除计算机可读存储介质之外的任何计算机可读介质,该介质可以通过连接至一个指令执行系统、装置或设备以实现通信、传播或传输供使用的程序。位于计算机可读信号介质上的程序代码可以通过任何合适的介质进行传播,包括无线电、电缆、光纤电缆、RF等,或任何上述介质的组合。

[0151] 本申请各部分操作所需的计算机程序代码可以用任意一种或以上程序设计语言编写,包括面向对象程序设计语言如Java、Scala、Smalltalk、Eiffel、JADE、Emerald、C++、C#、VB.NET、Python等,常规程序化程序设计语言如C程序设计语言、Visual Basic、Fortran 1703、Perl、COBOL 1702、PHP、ABAP,动态程序设计语言如Python、Ruby和Groovy,或其他程序设计语言等。该程序代码可以完全在用户计算机上运行、或作为独立的软件包在用户计算机上运行、或部分在用户计算机上运行部分在远程计算机运行、或完全在远程计算机或服务器上运行。在后种情况下,远程计算机可以通过任何网络形式与用户计算机连接,比如局域网(LAN)或广域网(WAN),或连接至外部计算机(例如通过因特网),或在云计算环境中,或作为服务使用如软件即服务(SaaS)。

[0152] 此外,除非权利要求中明确说明,本申请所述处理元素和序列的顺序、数字字母的使用、或其他名称的使用,并非用于限定本申请流程和方法的顺序。尽管上述披露中通过各种示例讨论了一些目前认为有用的发明实施例,但应当理解的是,该类细节仅起到说明的目的,附加的权利要求并不仅限于披露的实施例,相反,权利要求旨在覆盖所有符合本申请实施例实质和范围的修正和等价组合。例如,虽然以上所描述的系统组件可以通过硬件设

备实现,但是也可以只通过软件的解决方案得以实现,如在现有的服务器或移动设备上安装所描述的系统。

[0153] 同理,应当注意的是,为了简化本申请披露的表述,从而帮助对一个或以上发明实施例的理解,前文对本申请实施例的描述中,有时会将多种特征归并至一个实施例、附图或对其的描述中。然而,本申请的该方法不应被解释为反映所声称的待扫描对象物质需要比每个权利要求中明确记载的更多特征的意图。实际上,实施例的特征要少于上述披露的单个实施例的全部特征。

[0154] 一些实施例中使用了描述成分、属性数量的数字,应当理解的是,此类用于实施例描述的数字,在一些示例中使用了修饰词“大约”、“近似”或“大体上”来修饰。除非另外说明,“大约”、“近似”或“大体上”表明所述数字允许有 $\pm 20\%$ 的变化。相应地,在一些实施例中,说明书和权利要求中使用的数值参数均为近似值,该近似值根据个别实施例所需特点可以发生改变。在一些实施例中,数值参数应考虑规定的有效数位并采用一般位数保留的方法。尽管本申请一些实施例中用于确认其范围广度的数值域和参数为近似值,在具体实施例中,此类数值的设定在可行范围内尽可能精确。

[0155] 本文中提及的所有专利、专利申请、专利申请公布和其他材料(如论文、书籍、说明书、出版物、记录、事物和/或类似的东西)均在此通过引用的方式全部并入本文以达到所有目的,与上述文件相关联的任何起诉文档记录、与本文件不一致或冲突的任何上述文件或迟早与本文件相关联的权利要求书的广泛范畴有限定作用的所有上述文件除外。举例来说,如果与任何合并材料相关联的术语的描述、定义和/或使用与本文件相关联的术语之间存在任何不一致或冲突,以本文件中的术语的描述、定义和/或用法为准。

[0156] 最后,应当理解的是,本申请中所述实施例仅用以说明本申请实施例的原则。其他的变形也可能属于本申请的范围。因此,作为示例而非限制,本申请实施例的替代配置可视为与本申请的教导一致。因此,本申请的实施例不限于精确地如所示和所述的那些。

100

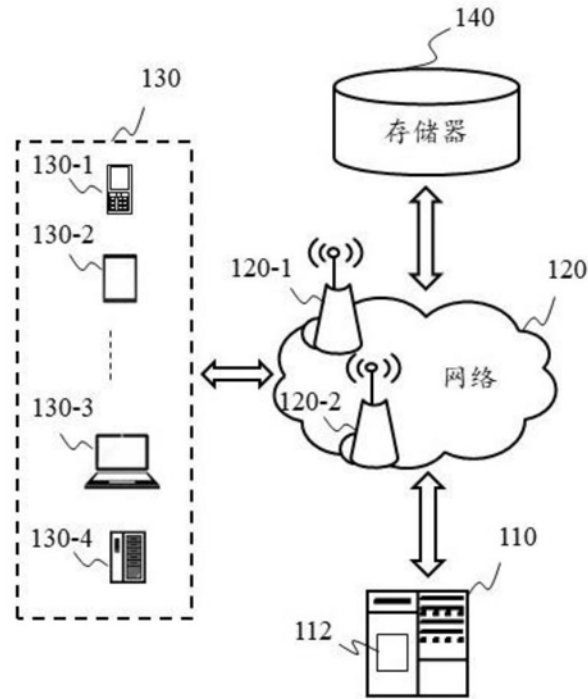


图1

200

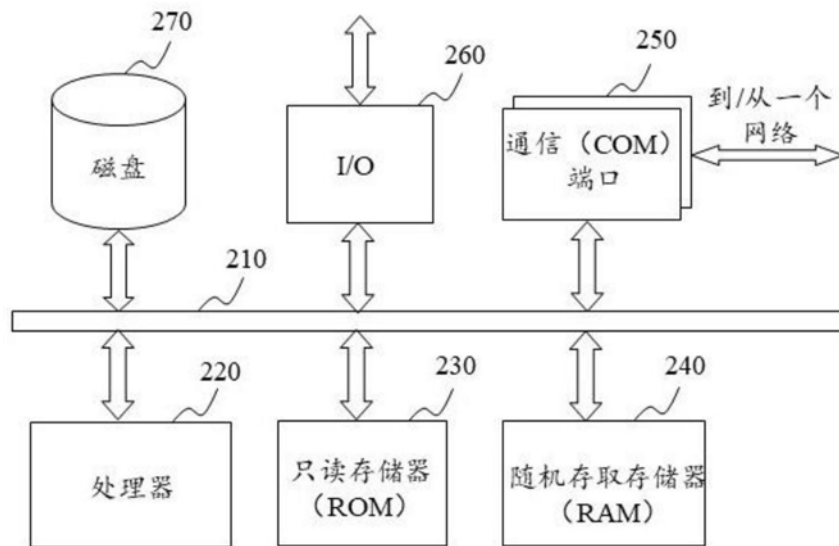


图2

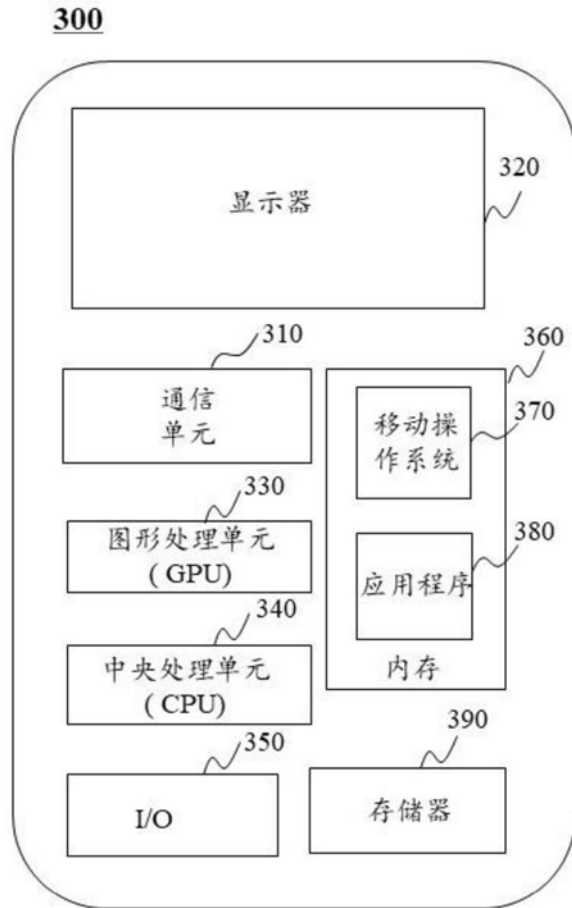


图3

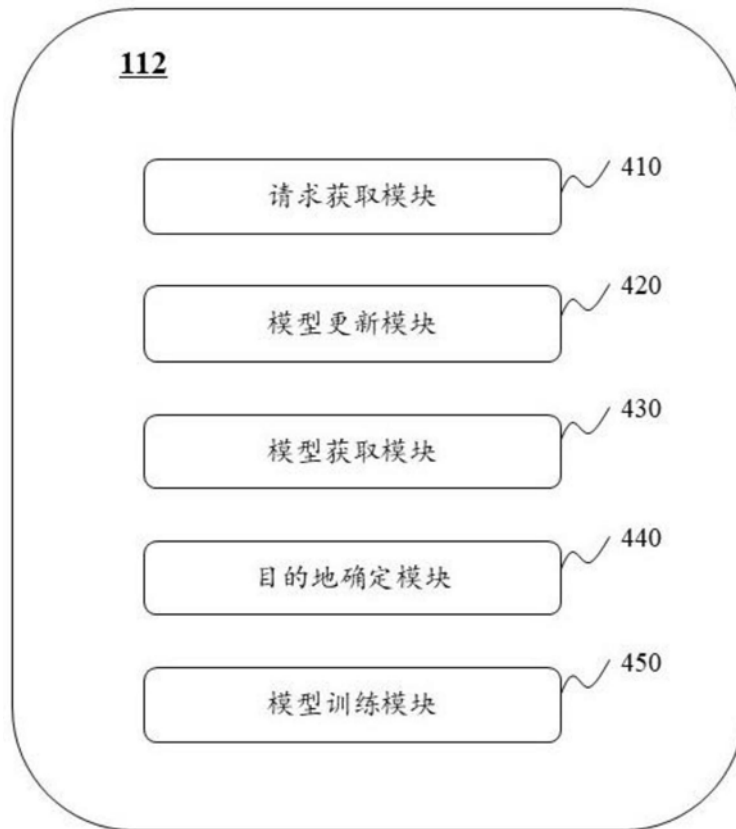


图4

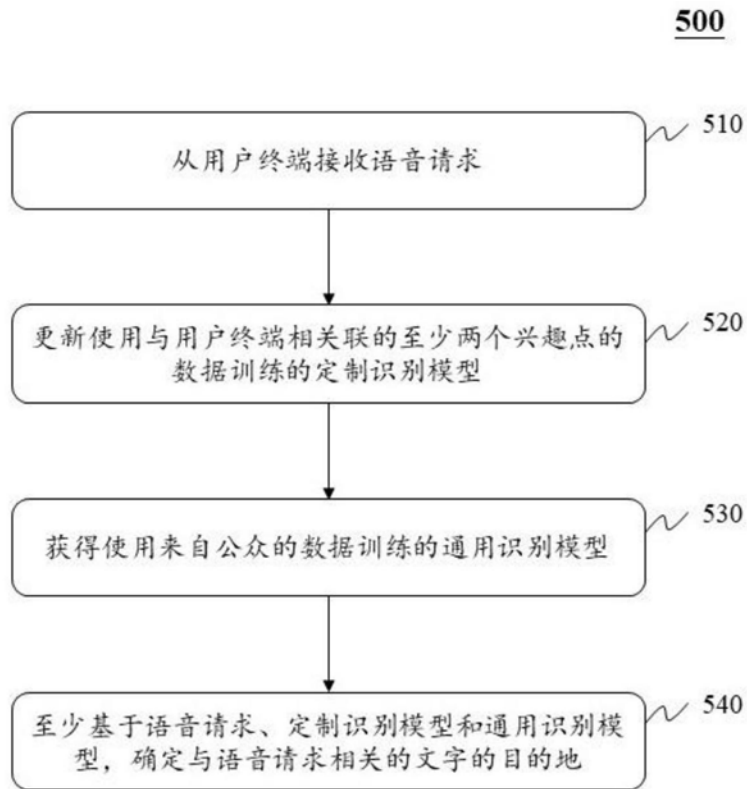


图5A

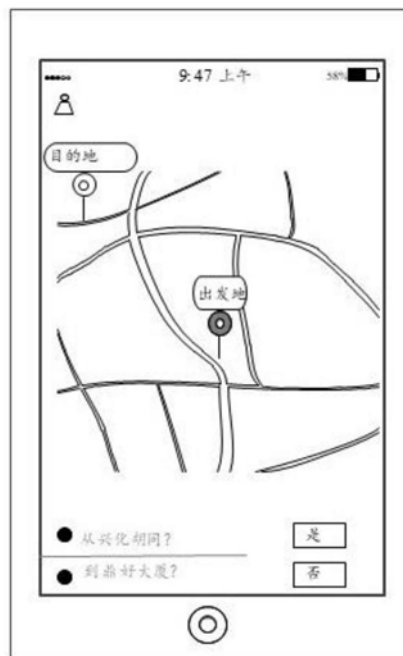


图5B

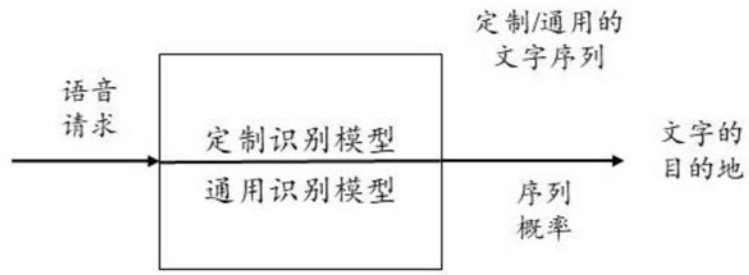


图6A

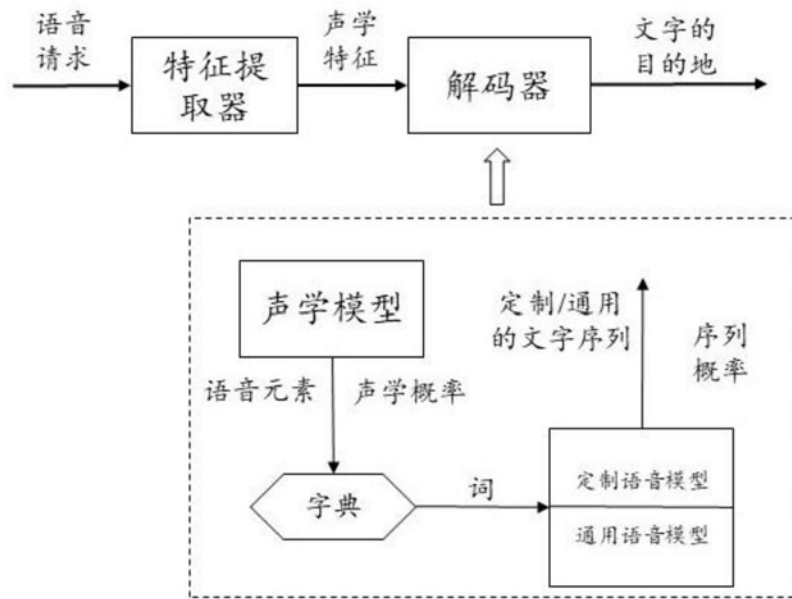


图6B

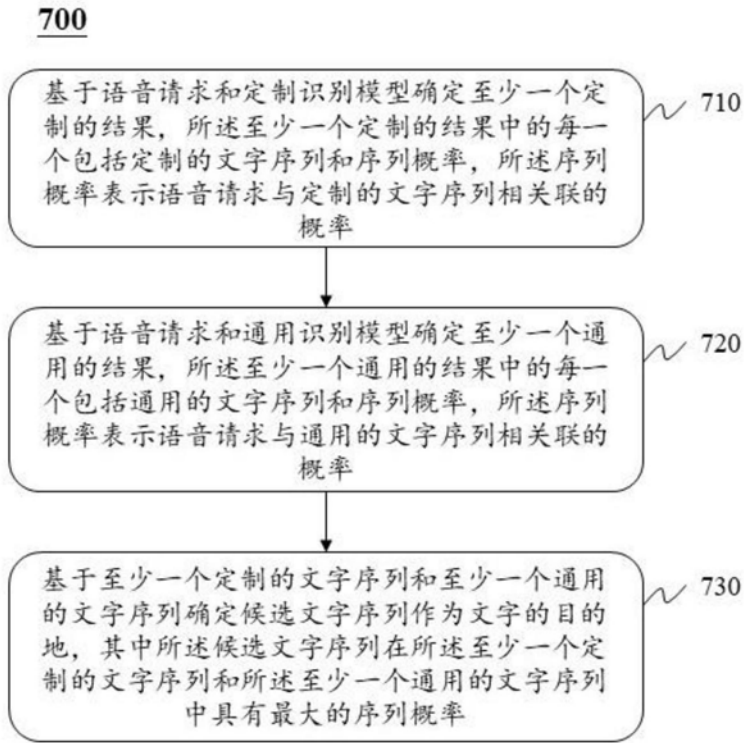


图7

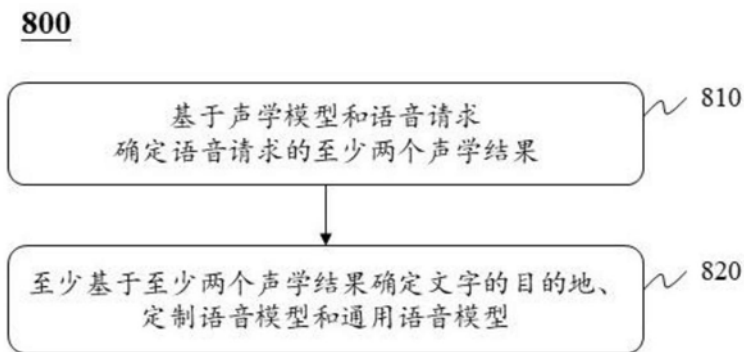


图8

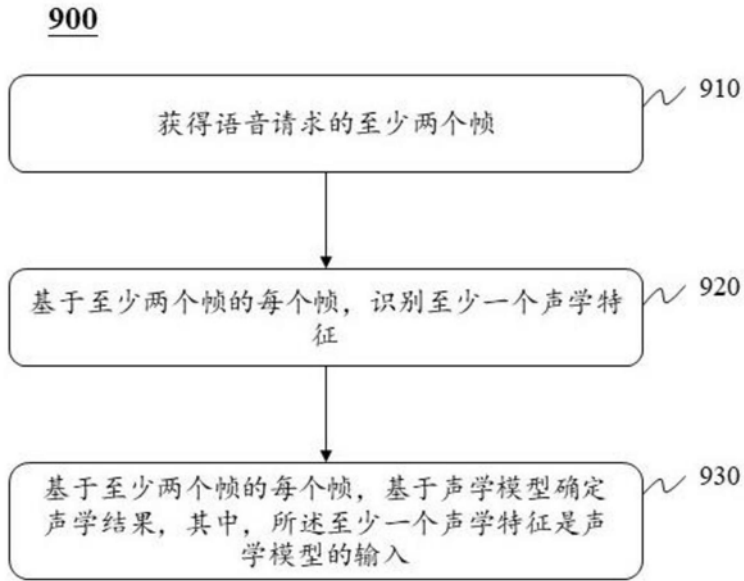


图9

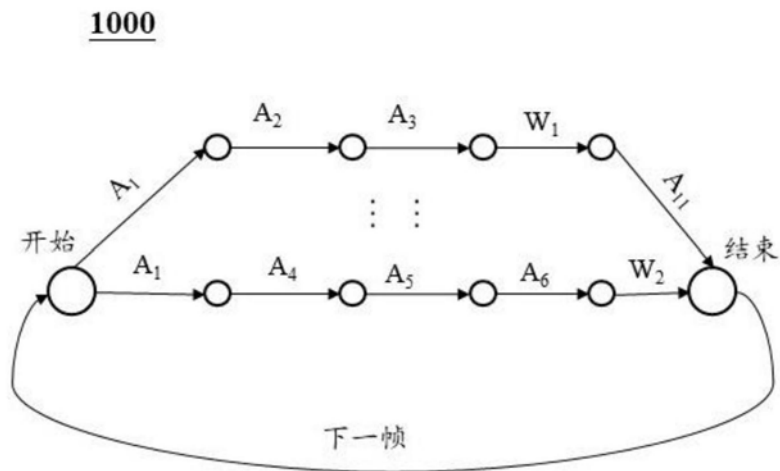


图10

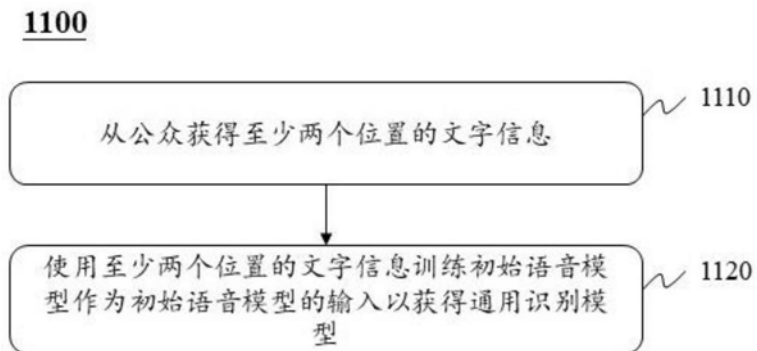


图11