(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112292675 A
(43)申请公布日 2021.01.29

(21)申请号 201880094970.9

(22)申请日 2018.05.01

(30)优先权数据
15/961,517 2018.04.24 US

(85)PCT国际申请进入国家阶段日
2020.12.23

(86)PCT国际申请的申请数据
PCT/US2018/030390 2018.05.01

(87)PCT国际申请的公布数据
WO2019/209351 EN 2019.10.31

(71)申请人 脸谱公司
地址 美国加利福尼亚州

(72)发明人 马丁·斯科特

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
代理人 韩倩倩 杨明钊

(51)Int.Cl.
G06F 16/00(2006.01)

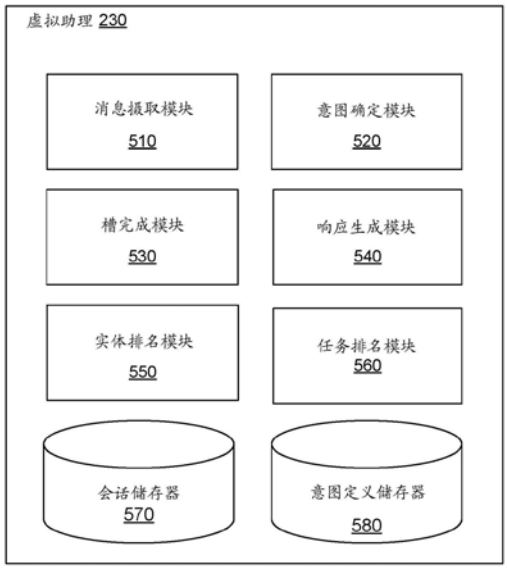
权利要求书3页 说明书14页 附图6页

(54)发明名称

利用实体和任务的显著性排名辅助计算机解译自然语言输入

(57)摘要

计算设备接收作为会话的一部分的消息。该消息包括从用户输入导出的自然语言内容。计算设备分析自然语言内容以确定对应于用户希望计算设备执行的给定任务的意图。该分析是基于任务的排名列表，该列表指示任务在会话中的当前相关性。从数据储存器检索包括实体槽的给定任务的定义，并且基于实体的排名列表将特定实体分配给该槽，该实体的排名列表指示实体在会话中的当前相关性。计算设备基于分配给实体槽的特定实体调用代理来执行给定任务。



CN 112292675 A

1.一种方法,包括:

在计算设备处接收消息,所述消息是会话的一部分并且包括从用户的输入导出的自然语言内容;

分析所述自然语言内容以确定意图,所述意图对应于用户期望所述计算设备执行的给定任务,所述分析基于任务的排名列表,所述任务的排名列表指示任务在所述会话中的当前相关性;

从数据储存器检索所述给定任务的定义,所述定义包括实体槽;

基于实体的排名列表,将特定实体分配给所述实体槽,所述实体的排名列表指示实体在所述会话中的当前相关性;以及

调用代理来执行所述给定任务,所述代理基于分配给所述实体槽的所述特定实体来执行所述给定任务。

2.根据权利要求1所述的方法,还包括基于所述消息更新所述任务的排名列表,更新所述任务的排名列表包括基于多个特征生成任务的排名分数,所述多个特征包括当前会话中任务的调用次数和当前会话中任务的最近度度量。

3.根据权利要求1所述的方法,其中,确定所述意图包括:

计算包括所述给定任务的多个任务的置信度值,每个置信度值指示相应任务对应于所述意图的可能性;以及

基于所述置信度值选择所述给定任务。

4.根据权利要求1所述的方法,其中,确定所述意图包括:

计算多个任务的置信度值,每个置信度值指示相应任务对应于所述意图的可能性;以及

响应于所述置信度值都没有超过意图置信度阈值,向用户的计算设备发送请求,所述请求要求用户澄清所述意图。

5.根据权利要求1所述的方法,其中,所述实体槽具有槽类型,并且将所述特定实体分配给所述实体槽包括:

按照排名顺序考虑所述实体的排名列表中的实体;以及

选择具有与所述槽类型匹配的实体类型的排名最高的实体作为所述特定实体。

6.根据权利要求1所述的方法,所述方法还包括基于所述消息更新所述实体的排名列表,更新所述实体的排名列表包括基于多个特征生成实体的排名分数,所述多个特征包括在当前会话期间提到实体的次数和会话中实体的最近度度量。

7.根据权利要求6所述的方法,其中,所述多个特征还包括独立于当前会话的关于用户的信息,关于用户的所述信息指示实体与用户的相关性。

8.一种存储计算机程序指令的非暂时性计算机可读介质,所述计算机程序指令能够由处理器执行以执行包括以下项的操作:

在计算设备处接收消息,所述消息是会话的一部分并且包括从用户的输入导出的自然语言内容;

分析所述自然语言内容以确定意图,所述意图对应于用户期望所述计算设备执行的给定任务,所述分析基于任务的排名列表,所述任务的排名列表指示任务在所述会话中的当前相关性;

从数据储存器检索所述给定任务的定义,所述定义包括实体槽;

基于实体的排名列表,将特定实体分配给所述实体槽,所述实体的排名列表指示实体在所述会话中的当前相关性;以及

调用代理来执行所述给定任务,所述代理基于分配给所述实体槽的所述特定实体来执行所述给定任务。

9.根据权利要求8所述的非暂时性计算机可读介质,其中,所述操作还包括基于所述消息更新所述任务的排名列表,更新所述任务的排名列表包括基于多个特征生成任务的排名分数,所述多个特征包括当前会话中任务的调用次数和当前会话中任务的最近度度量。

10.根据权利要求8所述的非暂时性计算机可读介质,其中,确定所述意图包括:

计算包括所述给定任务的多个任务的置信度值,每个置信度值指示相应任务对应于所述意图的可能性;以及

基于所述置信度值选择所述给定任务。

11.根据权利要求8所述的非暂时性计算机可读介质,其中,确定所述意图包括:

计算多个任务的置信度值,每个置信度值指示相应任务对应于所述意图的可能性;以及

响应于所述置信度值都没有超过意图置信度阈值,向用户的计算设备发送请求,所述请求要求用户澄清所述意图。

12.根据权利要求8所述的非暂时性计算机可读介质,其中,所述实体槽具有槽类型,并且将所述特定实体分配给所述实体槽包括:

按照排名顺序考虑所述实体的排名列表中的实体;以及

选择具有与所述槽类型匹配的实体类型的排名最高的实体作为所述特定实体。

13.根据权利要求8所述的非暂时性计算机可读介质,其中,所述操作还包括基于所述消息更新所述实体的排名列表,更新所述实体的排名列表包括基于多个特征生成实体的排名分数,所述多个特征包括在当前会话期间提到实体的次数和会话中实体的最近度度量。

14.根据权利要求13所述的非暂时性计算机可读介质,其中,所述多个特征还包括独立于当前会话的关于用户的信息,关于用户的所述信息指示实体与用户的相关性。

15.一种计算机系统,包括:

一个或更多个处理器;以及

存储计算机程序指令的一个或更多个计算机可读介质,所述计算机程序指令在被执行时使得所述一个或更多个处理器执行操作,所述操作包括:

在计算设备处接收消息,所述消息是会话的一部分并且包括从用户的输入导出的自然语言内容;

分析所述自然语言内容以确定意图,所述意图对应于用户期望所述计算设备执行的给定任务,所述分析基于任务的排名列表,所述任务的排名列表指示任务在所述会话中的当前相关性;

从数据储存器检索所述给定任务的定义,所述定义包括实体槽;

基于实体的排名列表,将特定实体分配给所述实体槽,所述实体的排名列表指示实体在所述会话中的当前相关性;以及

调用代理来执行所述给定任务,所述代理基于分配给所述实体槽的所述特定实体来执

行所述给定任务。

16.根据权利要求15所述的计算机系统，其中，所述操作还包括基于所述消息更新所述任务的排名列表，更新所述任务的排名列表包括基于多个特征生成任务的排名分数，所述多个特征包括当前会话中任务的调用次数和当前会话中任务的最近度度量。

17.根据权利要求15所述的计算机系统，其中，确定所述意图包括：

计算包括所述给定任务的多个任务的置信度值，每个置信度值指示相应任务对应于所述意图的可能性；以及

基于所述置信度值选择所述给定任务。

18.根据权利要求15所述的计算机系统，其中，确定所述意图包括：

计算多个任务的置信度值，每个置信度值指示相应任务对应于所述意图的可能性；以及

响应于所述置信度值都没有超过意图置信度阈值，向用户的计算设备发送请求，所述请求要求用户澄清所述意图。

19.根据权利要求15所述的计算机系统，其中，所述实体槽具有槽类型，并且将所述特定实体分配给所述实体槽包括：

按照排名顺序考虑所述实体的排名列表中的实体；以及

选择具有与所述槽类型匹配的实体类型的排名最高的实体作为所述特定实体。

20.根据权利要求15所述的计算机系统，其中所述操作还包括基于所述消息更新所述实体的排名列表，更新所述实体的排名列表包括基于多个特征生成实体的排名分数，所述多个特征包括在当前会话期间提到实体的次数、会话中实体的最近度度量、以及独立于当前会话的关于用户的信息，关于用户的所述信息指示实体与用户的相关性。

利用实体和任务的显著性排名辅助计算机解译自然语言输入

[0001] 背景

[0002] 本公开大体涉及计算机用户界面，且特别地，涉及基于自然语言的计算机界面。

[0003] 随着计算能力的可用性提高，同时计算设备的尺寸和相关成本降低，计算设备的使用范围增加。最近增长的一个领域是虚拟助理，现在许多智能手机和独立设备上都有虚拟助理。虚拟助理与用户交互，帮助他们完成各种任务。例如，不需要用户打开web浏览器、键入对用户家乡天气的请求、选择搜索结果并滚动所得出的网页来确定当天的天气预报，而用户可以只需询问助理“旧金山明天的天气会是什么样?(what’s the weather going to be like tomorrow in San Francisco?)”，其意图是虚拟助理将提供天气预报的概述作为响应。

[0004] 然而，现有的计算机系统通常难以正确解译(interpret)自然语言输入。在人类可能要依靠上下文和“常识”来解译一个陈述的情形中，现有的计算机系统通常会错误地解译请求，或者根本不能将它们识别为请求。这导致了时间的浪费和失望，因为用户必须以更明确的方式重新构建请求。这甚至可能导致用户完全放弃自然语言接口，转而使用其他形式的输入。改进自然语言处理可以提高计算机系统的速度、准确性和易用性。这还可以实现计算机化虚拟助理以及人类与之交互的其他计算系统的扩展和改进的功能。

[0005] 概述

[0006] 计算设备(例如，虚拟助理)响应于从用户接收的消息来执行任务。计算设备在解译它接收的消息时使用动作和实体的排名列表。动作的排名列表指示排名的动作在会话(session)中的当前相关性。类似地，实体的排名列表指示排名的实体在会话中的当前相关性。

[0007] 在一个实施例中，计算设备接收作为会话的一部分的消息。该消息包括从用户的输入导出的自然语言内容。计算设备基于任务的排名列表来分析自然语言内容，以确定用户的意图，该意图对应于用户希望计算设备执行的任务。计算设备从数据储存器检索包括实体的槽(slot)的任务定义，并基于实体的排名列表将特定实体分配给该槽。计算设备调用代理来基于分配给槽的实体执行任务。

[0008] 附图简述

[0009] 图1是根据实施例的虚拟助理在其中进行操作的系统环境的框图。

[0010] 图2是根据实施例的图1中所示的在线系统的框图。

[0011] 图3是示出根据实施例的由虚拟助理系统使用的意图本体(ontology)的框图。

[0012] 图4是示出根据实施例的图3中所示本体中的意图节点的框图。

[0013] 图5是根据实施例的图2中所示的虚拟助理的框图。

[0014] 图6示出了根据实施例的适用于图1的系统环境的计算机。

[0015] 图7是示出根据实施例的使用显著性排名(salience rankings)来解译自然语言输入的方法的流程图。

[0016] 附图仅为了说明的目的而描绘各种实施例。本领域中的技术人员从下面的讨论中将容易认识到本文示出的结构和方法的可选的实施例可以被采用而不偏离本文所述的原

理。

[0017] 详细描述

[0018] 综述

[0019] 当人类交流时，他们不会孤立地看待单独的语句。相反，语句是根据对话(conversation)的上下文，以及个体之间先前的对话和更广泛的社会经验来解译的。例如，考虑以下对话：

[0020] 人X:明天想在我家烧烤(BBQ)吗?

[0021] 人Y:天气会怎么样?

[0022] 人X:天气应该是晴朗的，大约70度。

[0023] 在对话中，人Y没有明确陈述时间或地点。即便如此，人X推断，人Y希望知道明天(先前陈述中针对烧烤提议的时间)人X所居住城镇(从先前讨论推断出的知识)的天气预报。人类交流依赖于这种语言捷径的程度给计算机解译自然语言输入带来了挑战，计算机传统上单独解译每个输入语句。然而，人类用户期望能够以类似于人类助理的方式与计算机化的虚拟助理对话。因此，使计算机能够自动并正确地解译这种语句的方法可以显著提高与虚拟助理交互的质量和效率，从而提高这种技术对用户的实用性。这种方法还可以提高其他基于计算机的技术(在该技术中，自然语言输入是所期望的)的实用性。

[0024] 在各种实施例中，用户与虚拟助理进行对话。对话可以是口头的，文本的，或者两者兼而有之。在某些情况下，可以包括其他形式的媒体，诸如图像和视频。虚拟助理维护一对显著性排名列表:一个用于实体，一个用于任务。实体排名和任务排名分别指示实体和任务与当前对话的当前相关性。虚拟助理基于显著性排名列表解译从用户接收的消息。因此，虚拟助理可以成功地解决对实体或任务的模糊和隐含的引用。在对话的每个回合之后(例如，当用户或虚拟助理发送了新消息时)，虚拟助理基于消息的内容更新显著性排名列表。

[0025] 系统架构

[0026] 图1是用于在线系统140(例如，社交网络系统)的系统环境100的框图。图1所示的系统环境100包括一个或更多个客户端设备110、网络120、一个或更多个第三方系统130和在线系统140。在替代配置中，系统环境100中可以包括不同和/或附加的部件。例如，可替代地，在线系统140可以是内容共享网络或向用户提供内容的另一种系统。

[0027] 客户端设备110是能够接收用户输入以及经由网络120发送和/或接收数据的一个或更多个计算设备。客户端设备110可以执行虚拟助理230或者与虚拟助理230交互。在一个实施例中，客户端设备110是常规的计算机系统，诸如台式计算机或膝上型计算机。替代地，客户端设备110可以是具有计算机功能的设备，诸如个人数字助理(PDA)、移动电话、智能手机、平板设备、或其他合适的设备。客户端设备110被配置成经由网络120进行通信。在一个实施例中，客户端设备110执行允许客户端设备110的用户与在线系统140交互的应用。例如，客户端设备110执行浏览器应用，以使得能够经由网络120在客户端设备110和在线系统140之间进行交互。在另一实施例中，客户端设备110通过运行在客户端设备110的本机(native)操作系统(诸如IOS®或ANDROID™)上的应用编程接口(API)来与在线系统140交互。

[0028] 客户端设备110被配置成使用有线和/或无线通信系统经由网络120进行通信，网络120可以包括局域网和/或广域网的任何组合。在一个实施例中，网络120使用标准通信技

术和/或协议。例如，网络120包括使用诸如以太网、802.11、全球微波接入互操作性(WiMAX)、3G、4G、码分多址(CDMA)、数字用户线路(DSL)等技术的通信链路。用于经由网络120进行通信的连网协议的示例包括多协议标签交换(MPLS)、传输控制协议/互联网协议(TCP/IP)、超文本传输协议(HTTP)、简单邮件传输协议(SMTP)和文件传输协议(FTP)。可以使用任何合适的格式(诸如超文本标记语言(HTML)或可扩展标记语言(XML))来表示通过网络120交换的数据。在一些实施例中，可以使用任何合适的技术或多项技术来对网络120的所有或一些通信链路进行加密。

[0029] 一个或更多个第三方系统130可以耦合到网络120，以与在线系统140进行通信，这将在下面结合图2进行进一步的描述。在一个实施例中，第三方系统130是应用提供商，其传送描述应用的信息以用于由客户端设备110执行，或者向客户端设备传送数据以供在客户端设备上执行的应用使用。在其他实施例中，第三方系统130经由客户端设备110提供用于呈现的内容或其他信息。第三方系统130还可以向在线系统140传送信息，诸如广告、内容或关于第三方系统130提供的应用的信息。

[0030] 图2是根据实施例的在线系统140的架构的框图。图2所示的在线系统140包括用户简档储存器205、内容储存器210、动作记录器215、动作日志220、边储存器225、虚拟助理230、知识储存器235和web服务器240。在其他实施例中，在线系统140可以包括用于各种应用的附加的、更少的或不同的部件。例如，一些实施例可以省略知识储存器235。没有示出常规部件(诸如网络接口、安全功能、负载平衡器、故障转移服务器、管理和网络操作控制台等)，以便不使系统架构的细节模糊。

[0031] 在线系统140的每个用户与被存储在用户简档储存器205中的用户简档相关联。用户简档包括由用户明确共享的关于用户的声明性信息，并且还可以包括由在线系统140推断的简档信息。在一个实施例中，用户简档包括多个数据字段，每个数据字段描述相应在线系统用户的一个或更多个属性。存储在用户简档中的信息的示例包括生物识别信息、人口统计信息和其他类型的描述性信息(诸如工作经历、教育历史、性别、爱好或偏好、位置等)。用户简档还可以存储用户提供的其他信息，例如图像或视频。在某些实施例中，可以用标识图像中显示的在线系统用户的信息来对用户的图像进行标记，其中标识用户被标记的图像的信息存储在用户的用户简档中。用户简档储存器205中的用户简档还可以维护对相应用户对于内容储存器210中的内容项执行的并存储在动作日志220中的动作的引用。

[0032] 虽然用户简档储存器205中的用户简档经常与个人相关联，允许个人经由在线系统140彼此交互，但是也可以为诸如企业或组织的实体而存储用户简档。这允许实体在在线系统140上建立存在，用于与其他在线系统用户关连并交换内容。实体可以使用与该实体的用户简档相关联的品牌页面来发布关于其自身、关于其产品的信息或者向在线系统140的用户提供其他信息。在线系统140的其他用户可以关连到品牌页面以接收被发布到品牌页面的信息或者从品牌页面接收信息。与品牌页面相关联的用户简档可以包括关于实体自身的信息，以向用户提供关于实体的背景或信息数据。

[0033] 内容储存器210存储对象，这些对象各自表示各种类型的内容。对象表示的内容的示例包括页面帖子、状态更新、照片、视频、链接、共享的内容项、游戏应用成就、在本地企业处的签到(check-in)事件、品牌页面或任何其他类型的内容。在线系统用户可以创建由内容储存器210存储的对象，诸如状态更新、被用户标记为与在线系统140中的其他对象相关

联的照片、事件、群组或应用。在一些实施例中，从第三方应用或独立于在线系统140的第三方应用接收对象。在一个实施例中，内容储存器210中的对象表示单条内容(single piece of content)或内容“项(item)”。因此，通过经由各种通信渠道将文本和各种类型的媒体的内容项发布到在线系统140来鼓励在线系统用户与彼此进行通信。这增加了用户与彼此交互的量，并增加了用户在在线系统140内交互的频率。

[0034] 动作记录器215接收关于在在线系统140内部和/或外部的用户动作的通信，用关于用户动作的信息填充动作日志220。动作的示例包括添加与另一用户的关连、向另一用户发送消息、上传图像、从另一用户读取消息、查看与另一用户相关联的内容以及参加由另一用户发布的事件。另外，许多动作可能涉及一个对象和一个或更多个特定用户，因此这些动作也与特定用户相关联并存储在动作日志220中。

[0035] 动作日志220可以由在线系统140使用以跟踪在线系统上的用户动作以及向在线系统传送信息的第三方系统130上的动作。用户可以与在线系统140上的各种对象交互，并且描述这些交互的信息存储在动作日志220中。与对象的交互的示例包括：评论帖子、共享链接、经由客户端设备110在物理位置签到、访问内容项以及任何其他合适的交互。与可以包括在动作日志220中的、在线系统140上的对象的交互的附加示例包括：评论相册、与用户通信、建立与对象的关连、加入事件、加入群组、创建事件、授权应用、使用应用、表达对对象的偏好(“点赞(liking)”对象)以及参与交易。另外，动作日志220可以记录用户与在线系统140上的广告的交互以及与在线系统上运行的其他应用的交互。在一些实施例中，来自动作日志220的数据用于推断用户的兴趣或偏好，增强用户的用户简档中包括的兴趣，并允许对用户偏好的更完整的理解。

[0036] 动作日志220还可以存储在第三方系统130(例如外部网站)上得到的用户动作并传送给在线系统。例如，电子商务网站可以通过使电子商务网站能够识别在线系统的用户的社交插件来辨识在线系统140的用户。因为在线系统140的用户是唯一可识别的，所以(诸如在前面的示例中的)电子商务网站可以将关于用户在在线系统外部的动作的信息传送给在线系统，用于与用户进行关联。因此，动作日志220可以记录关于用户在第三方系统130上执行的动作的信息，包括网页浏览历史、参与的广告、完成的购买以及来自购物和购买的其他模式。另外，用户经由与第三方系统130相关联的应用来执行的且在客户端设备110上执行的动作可以由该应用传送给动作记录器215，用于在动作日志220中进行记录和与用户进行关联。

[0037] 在一个实施例中，边储存器225将描述在用户与在线系统140上的其他对象之间的关连的信息存储为边。某些边可以由用户定义，允许用户指定他们与其他用户的关系。例如，用户可以生成与其他用户的边，这些边并行于用户的现实生活关系(诸如朋友、同事、伙伴等)。当用户与在线系统140中的对象交互(诸如表达对在线系统上的页面的兴趣、与在线系统的其他用户共享链接、以及对由在线系统的其他用户制作的帖子进行评论)时，生成其他边。边可以关连社交网络中的两个关连用户，也可以将用户与系统中的一个对象关连。在一个实施例中，节点和边形成复杂的关连社交网络，以指示用户彼此之间如何相关或关连(例如，一个用户接受来自另一个用户的朋友请求以成为社交网络中的关连)，以及由于用户以某种方式与对象交互(例如，“点赞”页面对象、加入事件对象或群组对象等)，用户如何关连到该对象。对象也可以基于对象相关或它们之间具有某种交互来彼此关连。

[0038] 边可以包括各种特征，每个特征表示用户之间的交互、用户和对象之间的交互、或者对象之间的交互的特性。例如，边中包括的特征描述了两个用户之间的交互的速率、两个用户最近如何与彼此进行交互、一个用户检索关于对象的信息的速率或量、或者用户发布的关于对象的评论的数量和类型。这些特征还可以表示描述特定对象或用户的信息。例如，特征可以表示用户对特定主题的兴趣程度、用户登录在线系统140的速率、或者描述关于用户的人口统计信息的信息。每个特征可以与源对象或用户、目标对象或用户以及特征值相关联。特征可以被指定为基于描述源对象或用户、目标对象或用户、或者源对象或用户与目标对象或用户之间的交互的值的表达式；因此，边可以被表示为一个或更多个特征表达式。

[0039] 边储存器225还存储关于边的信息，诸如对对象、兴趣和其他用户的亲和力分数(affinity score)。在线系统140可以随时间计算亲和力分数或“亲和力”，以基于用户执行的动作来近似用户对在线系统中的另一用户或对象的兴趣。在线系统140可以随时间计算用户的亲和力，以基于用户执行的动作来近似用户对在线系统中的另一用户、对象或主题的兴趣。在一个实施例中，用户和特定对象之间的多个交互可以作为单条边存储在边储存器225中。替代地，用户和特定对象之间的每次交互都被存储为单独的边。在一些实施例中，用户之间的关连可以存储在用户简档储存器205中，或者用户简档储存器205可以访问边储存器225以确定用户之间的关连。

[0040] 虚拟助理230经由用户的客户端设备110与用户交互，以提供任务帮助。用户经由客户端设备110处的用户输入向虚拟助理230提交请求。在一个实施例中，虚拟助理230在客户端设备110上的消息传递应用中作为联系人出现。在另一个实施例中，用户使用客户端设备110上的虚拟助理应用与虚拟助理230交互。替代地，虚拟助理230可以是运行在客户端设备110上的应用。

[0041] 虚拟助理230维护当前会话的实体排名和任务排名，它应用这些排名来解译从用户接收的消息。当前会话可以是当前对话(例如，自大于阈值长度(诸如五分钟)的、没有任何消息的上一时间段以来发送的任何消息)、当前日历日、最近二十四小时、最近一周或任何其他时间段。当前会话甚至可以实际上是无限的，其包括用户和虚拟助理之间交换的在系统环境100中仍然可用的所有消息。用户还能够定义会话，诸如通过向虚拟助理230发送“开始新会话”消息。在一些实施例中，什么被认为是当前会话可以取决于当前对话中涉及的消息类型。例如，针对口头对话的会话可能限于当前对话，而基于文本的对话可能认为会话是上周、上月等内交换的所有消息。

[0042] 在一个实施例中，实体排名是在当前会话中已经被提到的实体的排序列表。实体可以是用户(或虚拟助理230)在消息中提到的任何有形事物或无形概念。示例包括人物、位置、日期、时间、对象、事件等。虚拟助理230基于指示实体在会话内的当前相关性的各种信号来对列表中的实体进行排名。

[0043] 在一个实施例中，任务排名是用户已经在当前会话中要求虚拟助理230执行的任务的排序列表。列表中的每个任务对应于用户在向虚拟助理230发送消息时可能想要虚拟助理230执行的一些动作。像实体一样，列表中的任务基于指示会话内当前相关性的信号进行排名。参考图5，更详细地描述了虚拟助理230的各种实施例和用于确定实体和任务排名的信号，以及如何使用这些排名来更有效地为用户意图服务。

[0044] 图3示出了用户意图本体(ontology)300的一个实施例。当用户向虚拟助理230发

送消息时，用户想要虚拟助理执行特定任务。虚拟助理230分析该消息以确定用户针对该消息所具有的意图。换句话说，消息的意图对应于虚拟助理要执行的任务。通常，在意图(用户想要虚拟助理230做什么)和任务(虚拟助理做什么)之间存在一一对应关系。因此，任务的排名列表可以帮助虚拟助理230确定用户的意图。例如，与排名低的任务相比，用户更有可能希望虚拟助理230执行排名高的任务(即，当前在会话内非常相关的任务)。

[0045] 在图3所示的实施例中，意图被划分到本体300中的域(domain)中。本体300中的节点可以由数据对象(例如，存储在意图定义储存器580中)表示。在所示的实施例中，本体300具有根节点310，其用作整个本体的引用。在本体300内有几个域节点320，标记为D_1、D_2、…、D_n。域节点320对应于用户可能想要虚拟助理230执行的任务的一般类。例如，第一域节点320可以用于运输服务，第二域节点可以用于通信服务等。通过将意图排列到域中，可以提高虚拟助理230的效率。本体300的分层性质意味着虚拟助理230可以首先考虑与对应于最近执行的任务的意图在同一类中的意图。

[0046] 每个域节点320对应于被标记为I_{11}、I_{12}等的一组意图节点330。不同的域节点320可以对应于不同数量的意图节点330。每个意图节点330是定义如何执行相应任务的数据对象。例如，用于组织运输服务的域节点320可以链接到用于预订出租车、查询公共汽车时刻表、检查实时航班到达时间等的意图节点330。

[0047] 图4示出了意图节点330的一个实施例。意图节点330包括节点ID 410、一个或更多个槽420以及一个或更多个代理430。在所示的实施例中，槽420包括第一槽422至第N槽424，并且代理430包括第一代理432至第N代理434。然而，在一些实施例中，意图节点330可以包括单个槽420和/或代理430。意图节点330可以被视为定义虚拟助理230可以如何执行对应于该意图的任务的任务定义。在其他实施例中，意图节点可以包括不同的或附加的元素。

[0048] 节点ID 410标识本体300中的节点。节点ID 410可以是全局唯一标识号、对应于域节点320的标识符和名称中的标识号、具有节点名称的文本串等。

[0049] 槽420是虚拟助理230基于与用户的交互和/或其他可用信息来分配实体的变量。在一个实施例中，每个槽具有对应于将填充它的实体类型的类型。槽类型包括人物、位置、日期、时间、对象、事件等。槽可以具有子类型或范围重叠的类型(例如，可能存在可以保存时间、日、周、月等的一般“时间段”槽类型，以及针对时间、日、周、月等的单独槽类型)。举例来说，用于获取天气预报的任务的意图节点330的槽420可能包括位置槽、日期(日期范围)槽、和预报粒度槽。当被填充时，位置槽指示用户期望预报的位置，日期槽指示用户期望预报的时间段，预报粒度槽指示预报应该有多详细(例如，每小时、每日等)。换句话说，如果任务被视为虚拟助理230要执行的函数，则槽420类似于该函数的参量。

[0050] 代理430完成意图节点330想要服务的任务。例如，如果意图节点330的任务是在餐馆预定，代理430可以是提交相关信息(日期、时间、用餐人数等)到餐馆的在线预订系统的应用和/或功能。在一个实施例中，代理(例如，第一代理432)被直接构建到意图节点330中。换句话说，代理是执行所需任务的功能或程序。在另一个实施例中，代理是可以执行预期任务的程序或功能的指针或其他引用。因此，当调用代理时，使用指针或引用将相关信息发送给代理。这可以允许代理430独立于虚拟助理230被更新和/或允许第三方提供代理来(例如，经由第三方服务器130)服务特定意图。

[0051] 再次参考图2，根据一个实施例，知识储存器235存储已知或假定为事实的信息。这

是虚拟助理230可以用来为从用户接收的请求提供附加上下文或生成响应的信息。与存储关于关连用户和可用作现实世界关连的代理的其他对象的信息的边储存器225相反，知识储存器235存储从现实世界建立的信息。例如，如果用户问“剧院现在正在播放什么科幻电影?(what sci-fi movies are playing in theaters right now?)”，则虚拟助理230可以从第三方系统130检索电影放映列表，并基于知识储存器235中的流派信息(genre information)来确定这些电影中的哪些是科幻电影。

[0052] web服务器240经由网络120将在线系统140链接到一个或更多个客户端设备110以及链接到一个或更多个第三方系统130。web服务器240提供网页以及其他内容，诸如JAVA®、FLASH®、XML等。web服务器240可以接收消息并在在线系统140和客户端设备110之间路由该消息(例如即时消息、排队消息(例如，电子邮件)、文本消息、短消息服务(SMS)消息或使用任何其他合适的消息传送技术发送的消息)。用户可以向web服务器240发送上传存储在内容储存器210中的信息(例如，图像或视频)的请求。另外，web服务器240可以提供应用编程接口(API)功能，以将数据直接发送到本机客户端设备操作系统，诸如IOS®、ANDROID™或BlackberryOS。

[0053] 示例虚拟助理

[0054] 图5示出了虚拟助理230的一个实施例。在所示实施例中，虚拟助理230包括消息摄取模块510、意图确定模块520、槽完成模块530、响应生成模块540、实体排名模块550、任务排名模块560、会话储存器570和任务定义储存器580。在其他实施例中，虚拟助理230包含不同的和/或附加的元素。另外，功能可以以不同于描述的方式分布在这些元素中。

[0055] 消息摄取模块510从用户的客户端设备110接收消息。消息包括从用户输入导出的内容。内容可以是自然语言内容。在一个实施例中，用户可以通过输入文本或用口头话语来创建消息。如果用户输入文本，消息可以是用户输入的文本的副本。如果用户发出口头话语，消息可以包括话语的音频记录。消息摄取模块510利用语音到文本算法将音频记录转换成文本。替代地，记录可以在被发送到虚拟助理230之前由客户端设备110转换成文本。消息摄取模块510还可以(例如，在会话储存器570中)存储消息的副本。替代地，消息可以在该过程的稍后阶段存储(例如，在已经生成响应之后)。

[0056] 消息摄取模块510解析消息以识别对实体的引用。在一个实施例中，消息摄取模块510使用句法处理来识别对应于实体和那些实体的类型的单词或单词组。消息摄取模块510还可以将消息中的单词与包括已知实体和该实体的一种或更多种类型的实体字典进行比较。例如，通过与城市列表进行比较，单词“旧金山(San Francisco)”可以被识别为位置类型实体，即旧金山市。如果消息包含对实体的模糊引用，则消息摄取模块510将它们标记为模糊。它可能仍然能够将类型分配给实体。例如，代词“他”和“她”指人型实体，但这个人的身份在这个阶段被视为是模糊的。类似地，单词“那里”是对位置类型实体的明确引用，但是位置的标识被视为是模糊的。

[0057] 如果该消息是新会话中的第一个消息(例如，它之前有新的会话请求，或者前一个消息在超过阈值时间量的过去，等等)，消息摄取模块510创建实体和任务的新的排名列表(例如，在会话储存器570中)。在一个实施例中，意图列表最初是空的，并且实体列表仅包括在消息中被明确引用的实体。实体可以最初以随机顺序排名，或者它们可以基于独立于当前会话的信息(诸如指示用户与一个或更多个实体之间的关连的用户简档数据)来排序。这

些关连可以指示这些实体与用户的相关性。例如，如果消息提到旧金山和纽约，并且用户简档指示用户居住在旧金山，则旧金山最初可能排名高于纽约。在另一个实施例中，可以基于独立于会话的信息(诸如用户的简档)，用其他实体和/或动作来播种(seed)初始排名。例如，用户在社交网络上点赞或以其他方式与之关连的实体可以被包括在实体的排名列表中。

[0058] 如果该消息不是会话中的第一条消息，可以添加不在实体列表中的任何被引用实体。在一个实施例中，不在排名列表上的被引用实体被添加在底部(即，具有最低的排名)。因此，当排名列表在消息被处理后被重新排名时(如下所述，参考实体排名模块550)，新引用的实体将在列表上，并且由于在最近的消息中被引用，排名将有可能上移。替代地，消息摄取模块510可以调用实体排名模块550，以在消息被进一步处理之前对实体进行排名。这意味着当前消息中对实体的明确引用可以用来帮助消除消息中其他引用的模糊性。

[0059] 意图确定模块520分析由虚拟助理230接收的消息，以基于任务的排名列表来确定用户的意图。在各种实施例中，意图确定模块520遍历任务的排名列表(例如，从排名最高的任务开始)，并基于对消息中所识别实体的引用来寻找匹配。为了确定任务是否匹配用户的意图，意图确定模块520(例如，从意图定义储存器580)检索对应于任务的意图节点430，并将节点的槽420的类型与消息中所识别实体的类型进行比较。

[0060] 在一个实施例中，意图确定模块520确定是否所有的槽420都可以由消息中(模糊地或明确地)引用的适当类型的实体来填充。如果是，则对应于该任务的意图被分配给该消息。如果不是，意图确定模块520考虑排名次高的任务以确定它是否匹配等等，直到找到匹配或者到达任务的排名列表的末尾。如果在没有找到匹配的情况下到达任务的排名列表的末尾，则意图确定模块520向用户的客户端设备110发送消息，请求澄清用户的意图。例如，消息可能会说“对不起，我不明白，请再问一遍”。

[0061] 在另一个实施例中，意图确定模块520还考虑是否可以基于实体的排名列表来推理性地填充一些或所有槽420。如果消息不包括填充槽420的适当类型的实体，但是在实体的排名列表中存在适当的实体，则意图确定模块520用该实体填充槽。在这种情况下，意图确定模块520生成置信度值，该置信度值指示正在考虑的任务与用户的意图相对应的可能性。如果置信度值超过意图置信度阈值(例如，85％)，则意图确定模块520可以将该意图分配给消息。

[0062] 替代地，意图确定模块520可以计算排名列表中一些(例如，前五个)或所有其他任务的置信度值。置信度值是通过组合任务的排名与所识别实体匹配对应于任务的意图的槽420的程度来计算的。例如，可以基于任务对基于消息的意图的适合度和每个所识别实体对意图中一个槽的适合度分数，为考虑中的每个意图分配合成的排名。这些实体/槽兼容性分数可以启发式地或通过机器学习的权重来确定。在任一情况下，可以考虑诸如槽和实体类型的特征，以及实体在排名列表中的位置。如果没有任务具有超过意图置信度阈值的置信度值，或者多于一个任务具有超过阈值的置信度值，则意图确定模块520向用户的客户端设备110发送消息请求澄清。该消息可以识别最可能的任务(例如，具有最高置信度值的任务或排名最高的任务)并请求确认。例如，消息可以读取为“你要我给你旧金山明天的天气预报吗?(Do you want me to get you the weather forecast for tomorrow in San Francisco?)”。

[0063] 在一些实施例中，意图确定模块520从当前意图将与对应于由虚拟助理230执行的前一个任务的意图在同一域中的假设开始(例如，如会话储存器570中所示)。意图确定模块520识别前一个任务的域(例如，通过向本体查询对应于前一个任务的节点ID 410)，并针对相同域中的那些意图过滤动作的排名列表。意图确定模块520首先考虑对应于该域中的意图的动作(按排名顺序)。如果在当前域中没有找到匹配，则意图确定模块520然后考虑来自与其他域中的意图相对应的排名列表的动作(同样，按排名顺序)。

[0064] 假设已经选择了任务，槽完成模块530使用在消息中识别的实体和实体的排名列表来填充任务定义(例如，如从意图定义储存器580中的相应意图节点330检索到的)的槽420。这可能包括解决消息中的模糊引用(例如，代词指谁?)和/或确定隐含的引用(例如，在说“天气怎么样?”时，用户希望知道他或她在前一个消息中提到的地方的天气，或者那是前一个任务的结果)。

[0065] 在一个实施例中，在消息包括模糊引用或没有合适引用的情形中，对于每个槽，槽完成模块530尝试基于实体的排名列表来填充槽。槽完成模块530遍历实体的排名列表(从最高的排名开始)，并用具有匹配槽类型的类型的第一实体填充槽。实体的类型可以存储在排名列表中，或者可以使用句法分析和/或实体字典来确定。例如，如果槽是一个位置，则选择列表中排名最高的位置。在另一个实施例中，槽完成模块530首先过滤实体的排名列表，以移除不属于该槽的适当类型的实体。这可以提高效率，因为不属于该槽的适当类型的排名高的实体不需要被考虑。例如，如果槽是一个人，则排名列表中任何不是人的实体都不需要被考虑。

[0066] 在一些实施例中，槽完成模块530生成置信度值，该置信度值指示排名列表中的一些(例如，排名前五的实体)或所有实体匹配槽的可能性。实体的置信度值可以通过组合实体的排名、从消息的句法分析生成的分数和/或独立于当前会话的信息(例如，用户简档数据)来计算。例如，在位置实体的情况下，置信度值可以从发送消息的用户的当前位置的地理邻域导出。在个人实体的情况下，置信度值可以从(例如，如社交图边中所指示的)用户与所讨论实体之间的关系导出。置信度值还可能受最近是何时提及所讨论实体以及在当前对话线程中多少次提到该实体的影响。这些值可以启发式地或通过从机器学习模型获得的权重来确定。如果没有实体具有超过实体置信度阈值(例如，85％)的置信度值，则槽完成模块530向用户的客户端设备110发送消息请求澄清。该消息可以识别最有可能的实体(例如，具有最高置信度值或排名最高的实体)，并请求确认。例如，消息可以读取“你指的是约翰·史密斯(John Smith)吗?”。如果两个或更多个意图具有超过阈值的置信度值，则槽完成模块530也可以向用户请求澄清。例如，在来自用户的消息提到“约翰(John)”的情况下，槽完成模块350可以发送读取为“你指的是约翰·史密斯还是约翰·琼斯?(Are you referring to John Smith or John Jones?)”的消息。

[0067] 响应生成模块540生成对从用户的客户端设备110接收的消息的响应。在各种实施例中，响应生成模块540使用由意图确定模块540确定的意图和由槽完成模块540确定的槽值来识别适当的代理并制定向该代理的请求。如前所述，意图本体300中的意图节点330可以包括一组一个或更多个代理430。在有多于一个代理可以服务于该意图的情况下，响应生成模块540可以基于各种考虑来选择一个代理，这些考虑包括用户的偏好和槽值。

[0068] 在一个实施例中，响应生成模块540在用户第一次表达该意图时提示用户从该组

代理430中选择一个代理。将来，响应生成模块540可以自动选择相同的代理，除非用户明确地覆盖先前的选择(例如，通过向虚拟助理230发送请求忘记先前的代理选择的消息)。替代地或附加地，响应生成模块540可以在选择代理时考虑特定的槽值。意图节点330中的一个代理对于服务与一个实体相关的意图可能是优选的，而另一个代理关于不同的实体可能是优选的。例如，第一代理对于在旧金山预定餐馆可能更好(或用户更喜欢)，而第二代理对于在纽约预定餐馆可能更好(或用户更喜欢)。

[0069] 不管如何选择代理，响应生成模块540使用(如槽完成模块530所确定的)槽的值作为输入来调用所选择的代理。在一个实施例中，假设槽值定义了可行的调用请求，则代理返回所请求的结果。例如，如果代理用于获取天气信息，它将在请求的时间返回所请求位置的天气预报等等。作为另一个示例，如果代理是用于预定餐馆，它将返回已经预定的确认(假设它是成功的)。相反，如果调用请求不可行(例如，因为虚拟助理230向一个槽分配了无效值)，或者如果由于某种其他原因(例如，该请求是针对餐馆预定的，并且没有可用的桌子)，则代理可以返回错误。

[0070] 响应生成模块540基于从代理接收的结果创建要发送到用户的客户端设备110的消息。响应生成可以使用自然语言生成技术，使得消息的内容以与人类将创建的内容相似或相同的方式来构建。如果代理成功完成了预期的任务，则消息可以包括结果的细节(例如，餐馆预定的确认、天气预报等)。相反，如果代理的调用失败，则消息可以向用户寻求关于用户希望虚拟助理230执行什么任务的澄清。替代地，被调用的代理可以将结果直接返回给用户的客户端设备110。

[0071] 实体排名模块550基于虚拟助理230接收的消息和/或在会话期间发送给用户的客户端设备110的响应来更新会话实体的排名列表。在一个实施例中，实体排名模块550在对话的每个回合之后(例如，在接收到每个消息并且发送每个响应之后)对列表中的实体进行排名。替代地，可以以其他间隔执行排名，诸如在处理了用户发送的每个消息之后。不管使用的间隔如何，实体排名模块550通过计算实体的排名分数并基于排名分数对实体排序来更新排名列表。

[0072] 在一个实施例中，实体排名模块550基于一组特征来计算列表中每个实体的排名分数。该组特征可以包括：在当前会话期间提到该实体的次数；会话中实体的最近度(recency)度量(例如，自上次提到该实体以来的回合数)；指示实体对用户的重要性的用户简档信息；实体的上下文突出性(例如，在过去24小时内在社交网站上提到该实体的次数、该实体与会话中最近提到的其他实体之间的关连，等等)；用户和实体之间在消息传递应用(其中实体是具有这种账户的类型)中的交互频率；以及在社交网络中用户与实体的亲密程度的度量(例如，在实体是另一用户的情况下，用户和实体在社交网站上共同拥有多少朋友和/或兴趣)。例如，用户与在同一社交网络上拥有账户的个人实体的亲密程度的度量可以通过获得用户和所讨论实体共享的关连数，并用它除以用户的总关连数来导出。在其他实施例中，该组特征可以包括不同的或附加的特征。

[0073] 不管所使用的特定特征如何，实体排名模块550基于那些特征的值生成实体的排名分数。在一个实施例中，一组启发式预定的权重被分配给特征，其中分数是乘以相应权重的特征的总和。在另一个实施例中，使用机器学习模型来确定每个特征的权重。

[0074] 可以使用一组训练数据来训练机器学习模型。例如，一组人工操作员可以用分数

来标记用户和虚拟助理之间一组对话的训练，其中分数指示对话中各种实体的相对重要性。模型(例如，神经网络)被训练(例如，使用反向传播算法)以产生与人类分配的分数紧密匹配的训练集的分数。然后，可以通过将该模型应用于第二组对话并将结果与人工操作员为第二组对话分配的结果进行比较来验证该模型。

[0075] 不管用于生成排名分数的特定方法如何，实体排名模块550基于排名分数对实体列表进行排名。实体的排名列表提供了每个实体当前对对话的重要性的度量，较高的排名分数表示该实体可能更重要，反之亦然。排名高的实体(排名分数高)可能已经被多次提到(包括在最近的消息中)，并且用户可能以某种方式在社交网络上与该实体关连。相反，排名低的实体(排名分数低)可能在对话中的几个返回的消息中只被提到过一次，并且与用户没有其他明显的关连。

[0076] 任务排名模块560与实体排名模块550类似地操作，但是是对任务而不是对实体进行排名。在一个实施例中，任务排名模块560根据一组特征为每个任务生成排名分数，该组特征包括：当前会话中任务的调用次数；任务的最近度度量(例如，自上次任务调用以来的回合数)；所调用的最近n个任务中该任务的百分比(其中n是正整数，诸如10)；任务在用户地理区域中的突出性(例如，上周，任务在用户的邮政编码(ZIP code)中被调用了多少次)；任务在与用户相关联的其他用户(例如，用户在社交网络中的朋友、用户在即时消息传递服务中的联系人、与用户和虚拟助理230共享消息传递线程的用户等)中的突出性；以及任务与当前包括在实体列表中的实体(或排名前m的实体，其中m是正整数，例如10)类型的相关性度量。例如，任务与当前包括在实体列表中的实体类型的相关性可以通过为排名列表中类型与包括在该任务的定义中的槽的类型相匹配的每个实体递增任务的分数来确定。为给定实体递增的分数的量可以取决于该实体的排名。排名较高的实体可以通过应用乘数(例如，实体在排名列表中位置的倒数)来提供分数的较大增加。如先前参考实体排名模块550所描述的，每个信号对排名分数的贡献可以基于预定的启发式加权或使用机器学习方法来确定。

[0077] 不管它们是如何确定的，任务排名模块560基于排名分数对任务进行排名。因此，任务的排名列表提供了每个任务当前对对话有多重要的度量，较高的排名分数表示该任务可能比较重要，反之亦然。例如，如果用户询问特定城镇明天的天气，然后询问当晚在该城镇的各种酒店住宿的价格，则“获取酒店价格”任务可能排名较高，而“获取天气预报”任务可能排名较低。

[0078] 会话储存器570包括被配置成存储关于当前会话的信息的一个或更多个计算机可读介质。在一个实施例中，关于当前会话的信息包括实体和任务的排名列表以及由虚拟助理230在会话中发送和接收的每个消息的副本。会话储存器570还可以存储在先前会话期间交换的消息的副本。尽管会话储存器570被示为位于虚拟助理230内的单个实体，但是它可以位于别处和/或分布在多个计算设备上。例如，会话储存器570可以是存储在多个远程服务器上的分布式数据库。

[0079] 意图定义储存器580包括被配置成存储关于由虚拟助理230服务的意图的信息的一个或更多个计算机可读介质。在一个实施例中，关于意图的信息是意图层次结构300，如图3所示。像会话储存器570一样，意图定义储存器580可以远离虚拟助理和/或分布在多个计算设备上。另外，尽管图5中将会话储存器570和意图定义储存器580示为不同的部件，然

而相应的功能可以由单个数据库或其他类型的数据储存器提供。

[0080] 示例计算系统架构

[0081] 图6是示出适合用作在线系统140中的客户端设备110、第三方系统130或服务器的示例计算机600的高级框图。示例计算机600包括耦合到芯片组604的至少一个处理器602。芯片组604包括存储器控制器集线器620和输入/输出(I/O)控制器集线器622。存储器606和图形适配器612耦合到存储器控制器集线器620，而显示器618耦合到图形适配器612。存储设备608、键盘610、定点设备614和网络适配器616耦合到I/O控制器集线器622。计算机600的其他实施例具有不同的架构。

[0082] 在图6所示的示例中，存储设备608是非暂时性计算机可读存储介质，诸如硬盘驱动器、紧凑盘只读存储器(CD-ROM)、DVD或固态存储器设备。存储器606保存处理器602使用的指令和数据。定点设备614是鼠标、轨迹球、触摸屏或其他类型的定点设备，并且与键盘610(其可以是屏幕键盘)结合使用，以将数据输入到计算机系统600中。图形适配器612在显示器618上显示图像和其他信息。网络适配器616将计算机系统400耦合到一个或更多个计算机网络。

[0083] 图1、图2和图5的实体所使用的计算机的类型可以根据实施例和实体所需的处理能力而变化。例如，在线系统140可能包括分布式数据库系统，该分布式数据库系统包括一起工作以提供所描述功能的多个刀片服务器(blade server)。此外，计算机可以缺少一些上述部件，诸如键盘610、图形适配器612和显示器618。

[0084] 使用显著性排名解译人类输入的示例方法

[0085] 图7示出了使用显著性排名来解译自然语言输入的方法700的一个实施例。图7的步骤是从执行该方法的虚拟助理230的角度示出的。然而，一些或所有步骤可以由其他实体和/或部件来执行。另外，一些实施例可以并行执行这些步骤，以不同的顺序执行这些步骤，或者执行不同的步骤。例如，虚拟助理可以在生成对消息的响应之前或同时基于接收到的消息来更新显著性排名。

[0086] 在图7所示的实施例中，方法700开始于虚拟助理230从客户端设备110接收710消息。该消息包括从用户输入生成的内容。该内容可以包括(例如，在键盘上键入的)文本和/或(例如，由客户端设备110的麦克风记录的)用户说话的音频数据。在用户输入包括音频数据的情况下，虚拟助理230将其转换成文本(例如，利用语音到文本算法)。内容可以另外包括其他类型的媒体，例如图像和视频。

[0087] 虚拟助理230基于当前显著性排名来解译720消息。在一个实施例中，虚拟助理230基于任务的排名列表来确定消息的意图，并基于实体的排名列表解决对实体的任何模糊引用(代词、匹配多于一个实体的名称等)。如前所述，任务的排名列表指示每个任务与当前会话的相关性。类似地，实体的排名列表指示不同实体与当前会话的相关性。

[0088] 虚拟助理230调用730代理来完成对应于所识别的意图的任务。在一个实施例中，虚拟助理230(例如，从意图定义储存器380)检索意图的定义，该意图的定义限定了助理可以如何执行相应的任务。该定义包括由实体填充的一个或更多个槽。虚拟助理基于实体的排名列表填充槽，并调用730代理来完成任务。定义中可以包含或标识要调用的特定代理。代理完成任务并将结果返回给虚拟助理230。

[0089] 虚拟助理230生成740对消息的响应。响应包括代理返回的结果。替代地，代理可以

将结果直接返回给用户的客户端设备110。虚拟助理230提供750响应以呈现给用户。在一个实施例中，提供750响应包括向与用户相关联的客户端设备110(例如，从其接收消息的设备)发送响应。客户端设备110然后向用户显示调用代理的结果(例如，通过在用户使用其与虚拟助理230通信的应用中显示它)。

[0090] 虚拟助理230还基于所接收的消息(如在720所解译的)更新760显著性排名。在一个实施例中，虚拟助理230将消息添加到会话历史(例如，如会话储存器570中所存储的)，并对任务和实体的列表重新排名。在另一个实施例中，虚拟助理230仅调整在当前回合中执行的任务和消息中引用的任何实体的排名分数。这可能更有效，因为需要更少的计算，但也可能降低整体排名的准确性。在这样的实施例中，虚拟助理可以周期性地(例如，每十个回合)完全重新计算排名列表，以平衡效率和准确性。

[0091] 使用所描述的显著性排名的用户界面可以提供几个优点。其中一个优点是改进了意图级别的歧义消除。当用户询问可能是几样事物(例如，一本书、一部戏剧或一部电影，所有这些事物都具有相同的名称标题)之一的实体时，这种情况就会出现。任务排名可以帮助虚拟助理230正确地识别正确的身份，而不必向用户寻求明确的澄清。例如，如果用户最近询问了另外两本书，则最有可能的是，模糊请求的意图也是询问一本书，而不是一部戏剧或电影。因此，可以提高用户界面的效率，因为虚拟助理230可以在解译请求时犯更少的错误和/或可以不需要经常向用户请求澄清。

[0092] 另一个优点是提高了共指分辨率(coreference resolution)。这种情况出现在以下情况下：对话最近包括对两个实体的引用，然后用户发出包含代词的请求。实体排名可以提高虚拟助理230正确确定用户指的是哪个实体的能力。如果会话历史的总和指示正确的实体对整个对话更重要，即使最近的引用是针对不正确的实体，也有可能是这种情况。因此，如同意图级别的歧义消除一样，可以提高用户界面的效率。

[0093] 可以实现的类似优点是槽结转(slot carryover)。这是指这样一种情况，即用户一直在询问一件事情，并在返回到第一个主题之前简短地继续询问另一件事情。例如，用户可能会问几个关于餐馆的问题，然后问那天晚上是否会下雨，然后问“有没有无麸质的选项?(are there any gluten-free options?)”。人们很容易将最后一个问题理解为指的是餐馆，而不是相应位置的天气预报。实体排名可以使虚拟助理230能够正确地确定用户的意图，因为餐馆在会话历史中的整体突出性可以覆盖最近的天气引用。类似地，任务的排名列表可以使虚拟助理230能够成功地解译用户在短暂进入另一个主题之后返回到先前的意图。例如，如果用户请求各个滑雪场的可用酒店列表，然后询问其中一个滑雪场的天气，然后询问“Tahoe怎么样?”，虚拟助理230可以基于会话历史中酒店可用性任务的整体突出性，正确地将其解译为对酒店可用性的请求，而不是关于天气的请求。因此，可以提高用户界面的效率。用户可以以缩写的自然语言形式提交请求，并且虚拟助理可以需要更少的澄清和/或犯更少的错误。

[0094] 结论

[0095] 为了说明的目的提出了本公开的实施例的前述描述；它并不旨在是无遗漏的或将专利权利限制到所公开的精确形式。相关领域中的技术人员可以认识到，按照上面的公开，许多修改和变化是可能的。

[0096] 本描述的一些部分从对信息的操作的算法和符号表示方面描述了实施例。数据处

理领域的技术人员通常使用这些算法描述和表示来向本领域的其他技术人员有效地传达他们工作的实质。这些操作虽然在功能上、计算上或逻辑上进行了描述，但应理解为将由计算机程序或等效电路、微代码等来实现。此外，将操作的这些布置称为模块有时候也被证明是方便的而不失一般性。所描述的操作和它们的相关模块可以体现在软件、固件、硬件或其任何组合中。

[0097] 可以利用一个或更多个硬件或软件模块单独地或与其他设备组合地来执行或实现本文描述的任何步骤、操作或过程。在一个实施例中，利用包括包含计算机程序代码的计算机可读介质的计算机程序产品来实现软件模块，计算机程序代码可以由计算机处理器执行，用于执行所描述的任何或全部步骤、操作或过程。

[0098] 实施例也可以涉及用于执行本文的操作的装置。该装置可以被特别构造成用于所需的目的，和/或它可以包括由存储在计算机中的计算机程序选择性地激活或重新配置的通用计算设备。这种计算机程序可以存储在非暂时性的、有形的计算机可读存储介质中，或者适于存储电子指令的任何类型的介质中，这些介质可以耦合到计算机系统总线。此外，说明书中提到的任何计算系统可以包括单个处理器，或者可以是采用多处理器设计来提高计算能力的架构。

[0099] 实施例也可以涉及由本文所述的计算过程产生的产品。这样的产品可以包括由计算过程产生的信息，其中信息被存储在非暂时性的、有形的计算机可读介质上且可以包括计算机程序产品或本文所述的其他数据组合的任何实施例。

[0100] 最后，在说明书中使用的语言主要为了可读性和指导目的而被选择，并且它可以不被选择来描绘或限制专利权利。因此，意图是本专利权利的范围不受该详细描述限制，而是受在基于此的申请上所发布的任何权利要求限制。因此，实施方案的公开内容意图对本专利权利的范围是说明性的，而不是限制性的，在所附的权利要求中阐述了本专利权利的范围。

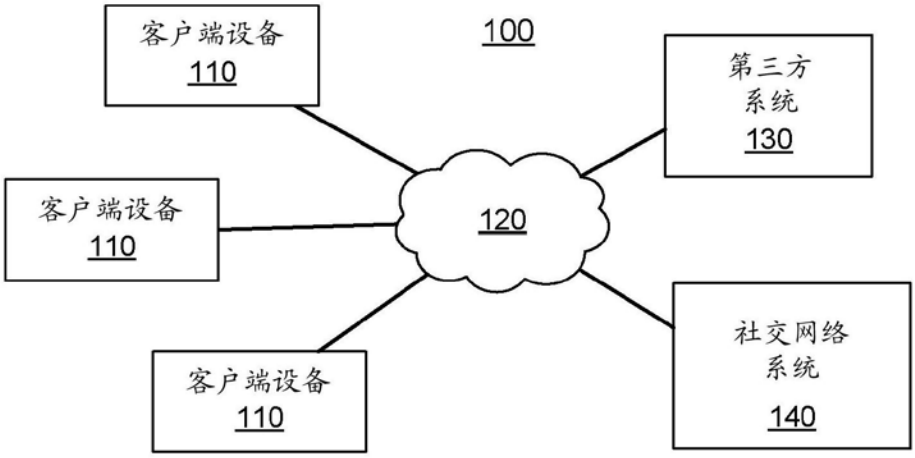

客户端设备
110
100
第三方
系统
130
客户端设备
110
120
客户端设备
110
社交网络
系统
140

图1

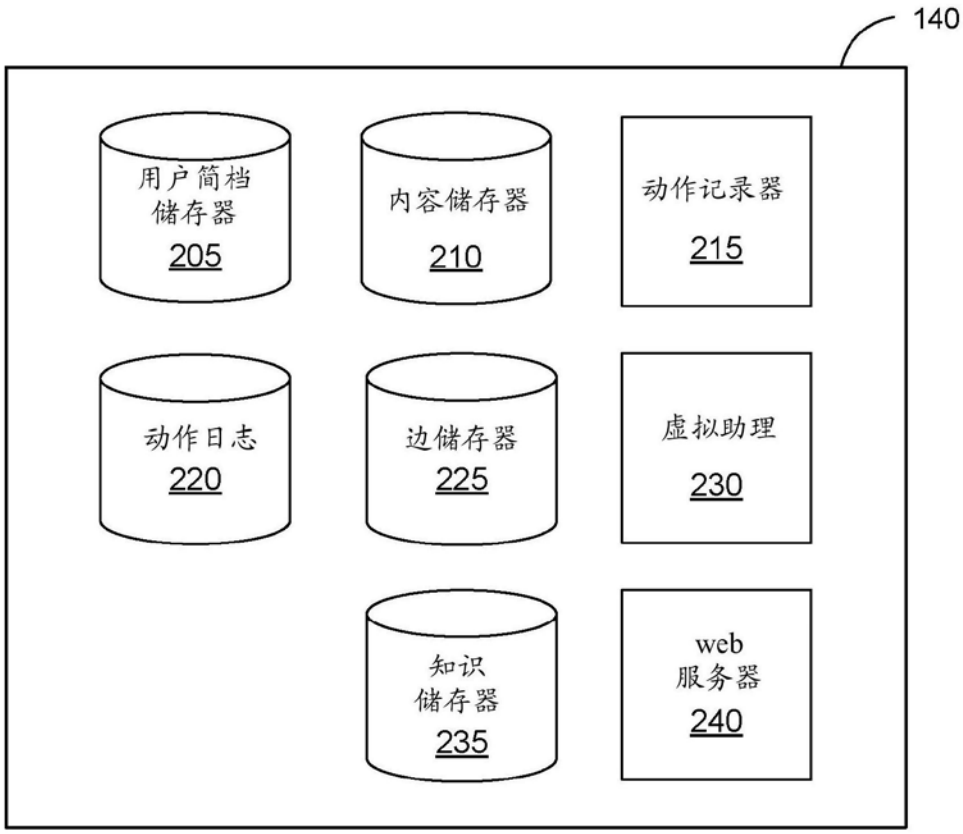

140
用户简档
储存器
205
内容储存器
210
动作记录器
215
动作日志
220
边储存器
225
虚拟助理
230
知识
储存器
235
web
服务器
240

图2

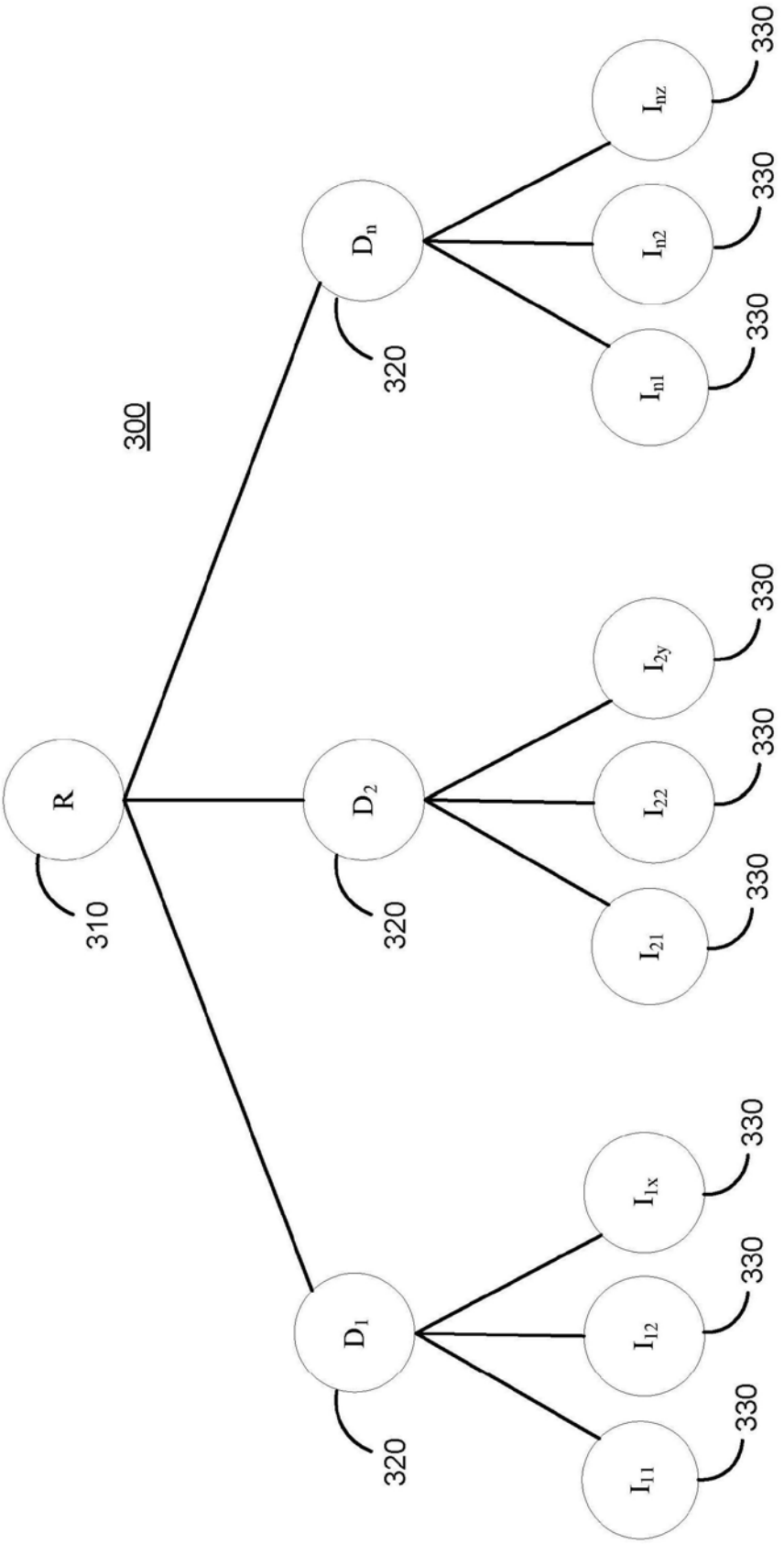


图3

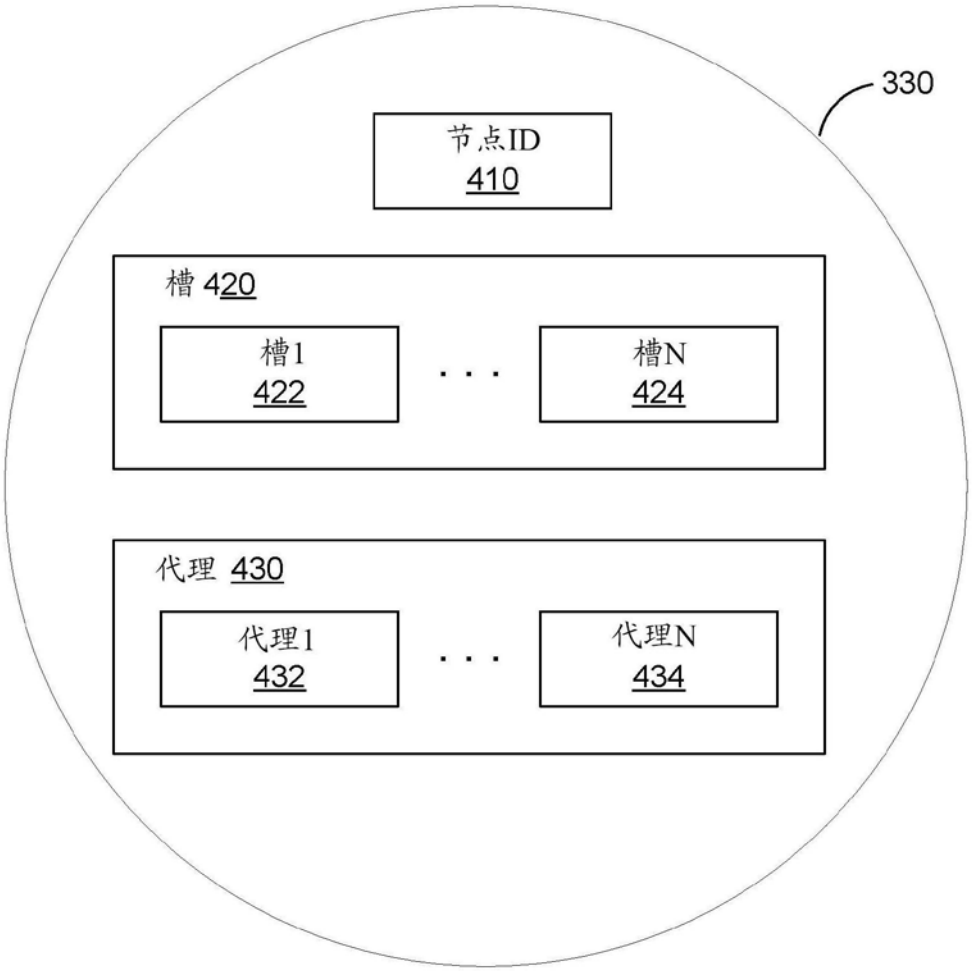


图4

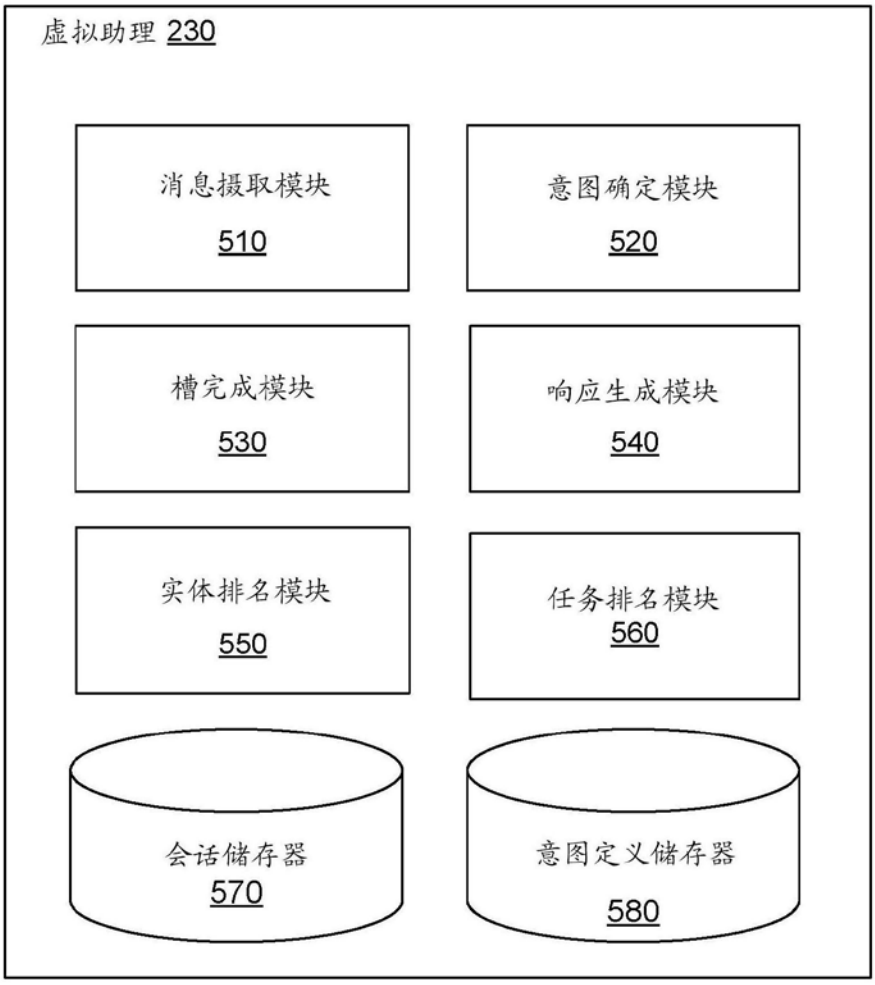

虚拟助理 230
消息摄取模块
510
意图确定模块
520
槽完成模块
530
响应生成模块
540
实体排名模块
550
任务排名模块
560
会话储存器
570
意图定义储存器
580

图5

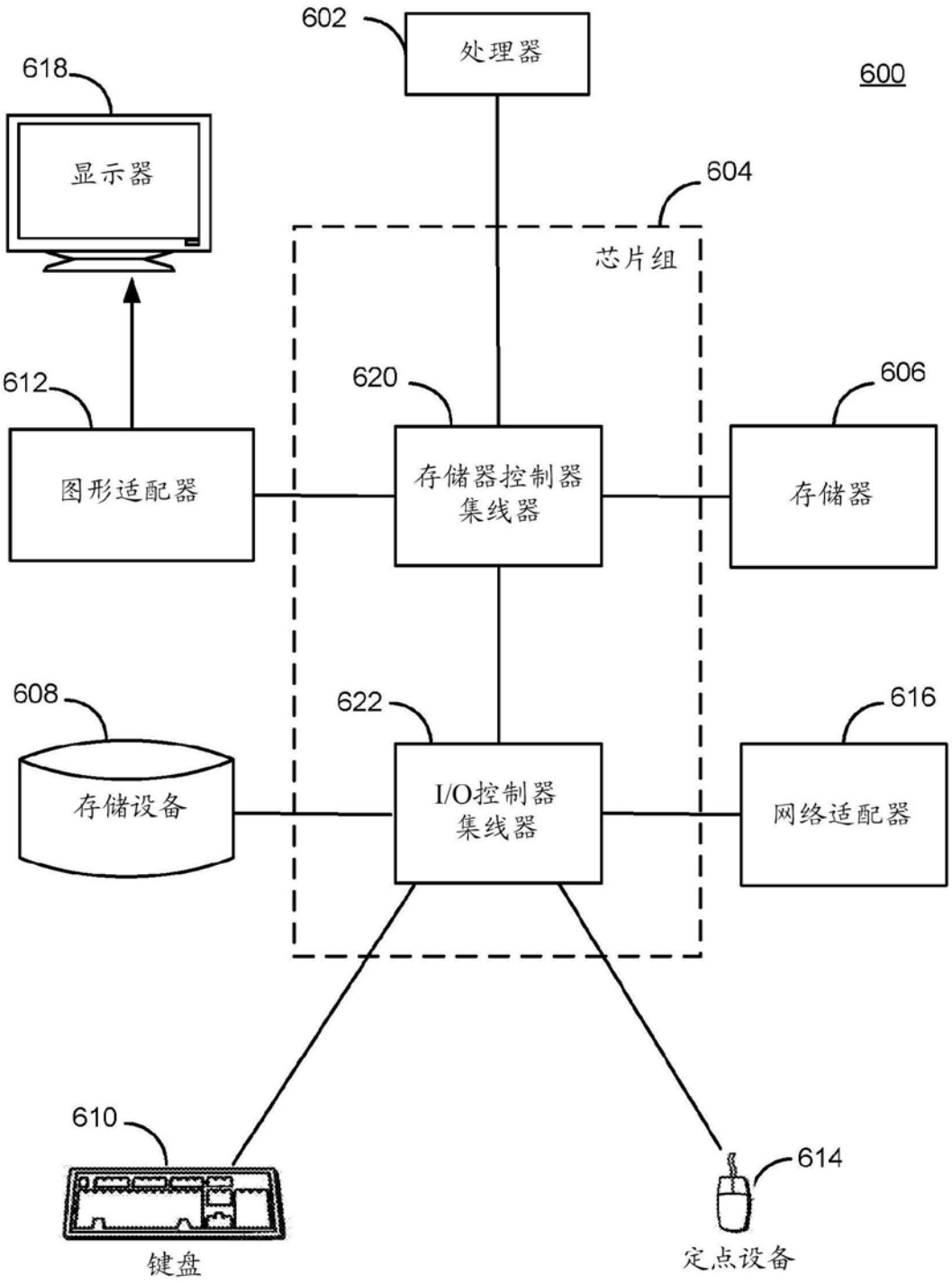

602
处理器
600
618
显示器
604
芯片组
612
620
606
图形适配器
存储器控制器
集线器
存储器
608
622
616
存储设备
I/O控制器
集线器
网络适配器
610
614
键盘
定点设备

图6

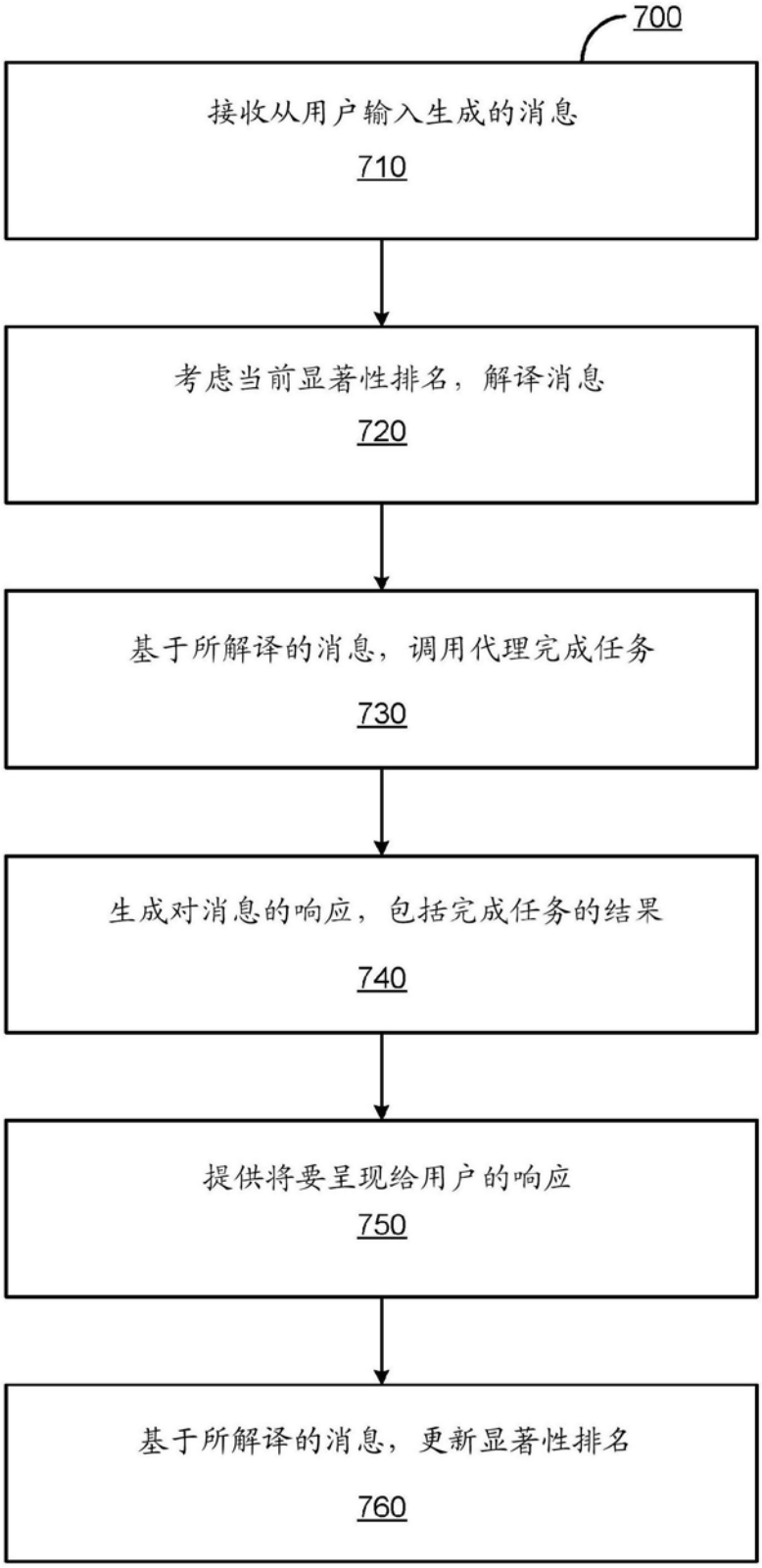

700
接收从用户输入生成的消息
710
考虑当前显著性排名，解译消息
720
基于所解译的消息，调用代理完成任务
730
生成对消息的响应，包括完成任务的结果
740
提供将要呈现给用户的响应
750
基于所解译的消息，更新显著性排名
760

图7