



(12) 发明专利申请

(10) 申请公布号 CN 112470160 A

(43) 申请公布日 2021.03.09

(21) 申请号 201980049312.2

(22) 申请日 2019.07.30

(30) 优先权数据

62/712,773 2018.07.31 US

16/404,012 2019.05.06 US

(85) PCT国际申请进入国家阶段日  
2021.01.22

(86) PCT国际申请的申请数据  
PCT/KR2019/009474 2019.07.30

(87) PCT国际申请的公布数据  
W02020/027540 EN 2020.02.06

(71) 申请人 三星电子株式会社  
地址 韩国京畿道

(72) 发明人 沈逸麟 曾翔宇 王昱 金红霞

(74) 专利代理机构 北京市立方律师事务所  
11330

代理人 谢玉斌

(51) Int.Cl.  
G06F 40/30 (2020.01)  
G10L 15/00 (2013.01)  
G10L 15/18 (2013.01)

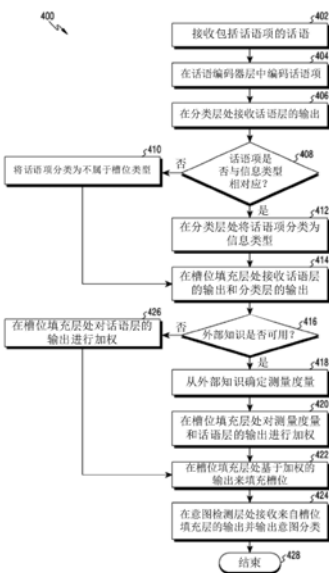
权利要求书2页 说明书18页 附图9页

(54) 发明名称

个性化自然语言理解的装置和方法

(57) 摘要

用于训练机器学习模型的电子设备包括至少一个存储器和耦接到该至少一个存储器的至少一个处理器。该至少一个处理器被配置为对模型的分层进行训练。为了对分层进行训练，该至少一个处理器被配置为通过分类层从话语编码器层接收一个或多个语言情景，并且通过分类层将话语的至少一部分分类为多个信息类型中的一个信息类型。至少一个处理器可以进一步被配置为对模型的槽位填充层和意图检测层进行联合训练。



1. 一种训练机器学习模型的方法,所述方法包括:  
对模型的分类层进行训练,包括:  
由所述分类层从话语编码器层接收一个或更多个语言情景;以及  
由所述分类层将话语的至少一部分分类为多个信息类型中的一个信息类型。
2. 根据权利要求1所述的方法,其中,将所述话语的至少一部分分类包括:将所述一个或更多个语言情景中的至少一个语言情景应用于在一个或更多个时间戳处的隐藏状态的加权总和。
3. 根据权利要求1所述的方法,所述方法进一步包括:  
对所述模型的槽位填充层和意图检测层进行联合训练,包括:  
由所述槽位填充层从所述话语编码器层接收所述一个或更多个语言情景,并且从所述分类层接收所述话语的已分类的至少一部分;以及  
基于所述一个或更多个语言情景以及所述话语的已分类的至少一部分,由所述槽位填充层填充一个或更多个槽位。
4. 根据权利要求3所述的方法,所述方法进一步包括:由所述意图检测层基于所述一个或更多个语言情景以及所述话语的已分类的至少一部分来确定意图。
5. 根据权利要求3所述的方法,其中,对所述槽位填充层和所述意图检测层进行联合训练进一步包括:  
由所述槽位填充层获取外部知识参数;  
基于所述外部知识参数确定至少一个测量度量;以及  
由所述槽位填充层对所述至少一个测量度量、所述话语的已分类的至少一部分以及所述一个或更多个语言情景中的至少一项进行加权,以确定槽位类型。
6. 根据权利要求5所述的方法,其中,所述外部知识参数是用户偏好数据;  
其中,所述至少一个测量度量是使用频率度量;并且  
其中,确定所述至少一个测量度量包括:基于所述用户偏好数据和所述话语的所述至少一部分来计算所述使用频率度量。
7. 一种用于自然语言理解的方法,所述方法包括:  
由分类层利用来自话语编码器层的一个或更多个语言情景将话语的至少一部分分类为多个信息类型中的一个信息类型;  
由槽位填充层从所述话语编码器层接收所述一个或更多个语言情景,并且从所述分类层接收所述话语的已分类的至少一部分;以及  
基于所述一个或更多个语言情景以及所述话语的已分类的至少一部分,由所述槽位填充层填充一个或更多个槽位。
8. 根据权利要求7所述的方法,其中,将所述话语的至少一部分分类包括:将所述一个或更多个语言情景中的至少一个语言情景应用于在一个或更多个时间戳处的隐藏状态的加权总和。
9. 根据权利要求7所述的方法,所述方法进一步包括:由意图检测层基于一个或更多个语言情景以及所述话语的已分类的至少一部分来确定意图。
10. 根据权利要求7所述的方法,所述方法进一步包括:  
由所述槽位填充层获取外部知识参数;

基于所述外部知识参数确定至少一个测量度量;以及

由所述槽位填充层对所述至少一个测量度量、所述话语的已分类的至少一部分以及所述一个或更多个语言情景中的至少一项进行加权,以确定槽位类型。

11. 根据权利要求10所述的方法,其中,所述外部知识参数是用户偏好数据;

其中,所述至少一个测量度量是使用频率度量;并且

其中,确定所述至少一个测量度量包括:基于所述用户偏好数据和所述话语的所述至少一部分来计算所述使用频率度量。

12. 根据权利要求5所述的方法,或者根据权利要求10所述的方法,其中,所述至少一个测量度量是通过在时间戳处的信息类型向量和所述外部知识参数的函数来确定的。

13. 根据权利要求5所述的方法,或者根据权利要求10所述的方法,其中,所述外部知识参数是位置数据;

其中,所述至少一个测量度量是距离度量;

其中,所述话语的已分类的至少一部分被分类为位置类型;并且

其中,确定所述至少一个测量度量包括:基于所述位置数据和所述话语的已分类的至少一部分所定义的位置来计算所述距离度量。

14. 一种用于训练机器学习模型的电子设备,所述电子设备包括:

至少一个存储器;以及

至少一个处理器,所述至少一个处理器耦接到所述至少一个存储器,其中,所述至少一个处理器被配置为对所述模型的分类层进行训练;

其中,为了训练所述模型的分类层,所述至少一个处理器被配置为实现权利要求1至13中任一项所述的方法。

15. 一种非暂时性计算机可读介质,包含用于操作包括至少一个处理器的电子设备的计算机程序,所述计算机程序包括计算机可读指令,所述计算机可读指令在由所述至少一个处理器执行时使所述电子设备对分类层进行训练;

其中,在被执行时使所述电子设备对所述分类层进行训练的所述计算机可读指令包括在被执行时使所述电子设备实现根据权利要求1至13中任一项所述的方法的计算机可读指令。

## 个性化自然语言理解的装置和方法

### 技术领域

[0001] 本公开总体上涉及机器学习系统。更具体地，本公开涉及用于个性化自然语言理解的装置和方法。

### 背景技术

[0002] 随着人工智能语音助手在日常生活中的出现，自然语言理解 (NLU) 领域正在不断发展。语义框架解析通常是 NLU 中引导对话系统的重要组成部分。语义框架解析通常包括两个主要任务，即意图检测和槽位填充。槽位填充包括从自然语言话语中提取语义成分。NLU 的深度学习模型通常不仅需要大规模的带注释的数据集 (大规模的带注释的语料库)，而且还需要很长的训练过程。由于 NLU 系统旨在了解所有类型的用户话语，因此语料库通常需要广泛涵盖所有类型的话语。然而，收集这种带注释的语料库是昂贵的并且需要大量的人工。为了开发新的域，必须生成新的话语数据集，并且训练新的语义框架解析模型需要很长时间。

### 发明内容

[0003] 技术问题

[0004] 本公开提供了用于个性化自然语言理解的系统和方法。

[0005] 问题的解决方案

[0006] 在第一实施例中，用于训练机器学习模型的方法包括对模型的分类层进行训练。对分类层进行训练包括由分类层从话语编码器层接收一个或更多个语言情景。对分类层进行训练还包括由分类层将话语的至少一部分分类为多个信息类型中的一个信息类型。

[0007] 在第二实施例中，用于训练机器学习模型的电子设备包括至少一个存储器和耦接到该至少一个存储器的至少一个处理器。该至少一个处理器被配置为对模型的分类层进行训练。为了对分类层进行训练，该至少一个处理器被配置为通过分类层从话语编码器层接收一个或更多个语言情景，并且通过分类层将话语的至少一部分分类为多个信息类型中的一个信息类型。

[0008] 在第三实施例中，一种非暂时性计算机可读介质实施一种用于操作包括至少一个处理器的电子设备的计算机程序。该计算机程序包括计算机可读指令，该计算机可读指令在由至少一个处理器执行时使电子设备对分类层进行训练。该计算机可读指令当被执行时，使得包括计算机可读指令的电子设备对分类层进行训练，该计算机可读指令当被执行时使得电子设备通过分类层从话语编码器层接收一个或更多个语言情景，并且通过分类层将话语的至少一部分分类为多个信息类型中的一个信息类型。

[0009] 在第四实施例中，用于自然语言理解的方法包括通过分类层利用来自话语编码器层的一个或更多个语言情景将话语的至少一部分分类为多个信息类型中的一个信息类型。该方法进一步包括由槽位填充层从话语编码器层接收一个或更多个语言情景，并且从分类层接收话语的已分类的至少一部分。该方法进一步包括：基于一个或更多个语言情景以及

话语的已分类的至少一部分,由槽位填充层填充一个或更多个槽位。

[0010] 在第五实施例中,用于自然语言理解的电子设备包括至少一个存储器和耦接到至少一个存储器的至少一个处理器。至少一个处理器被配置为执行第四实施例或其从属主张中的任一项的方法。在第六实施例中,非暂时性计算机可读介质包括计算机可读程序代码,该计算机可读程序代码当被执行时使至少一个处理器执行第四实施例或其从属主张中的任一项的方法。

[0011] 根据以下附图、说明书和权利要求书,其他技术特征对于本领域技术人员而言是显而易见的。

[0012] 本发明的有益效果

[0013] 根据本公开的各种实施例,可以具有以下技术效果:与采用大量历史数据作为训练模型的其他个性化机器学习模型不同,本公开不需要每个用户都具有大量的个性化训练数据。

[0014] 根据本公开的各种实施例,可以具有以下技术效果:本公开可以利用较少志愿者/测试用户的数据来平衡协作学习,并适应每个用户以在设备上获得个性化的NLU模型。

## 附图说明

[0015] 为了更加完整地理解本公开及其优点,现在参考结合附图的以下文字描述,在附图中相同的附图标记表示相同的组件:

[0016] 图1示出了根据本公开的各种实施例的示例网络配置;

[0017] 图2示出了根据本公开的各种实施例的示例机器学习架构;

[0018] 图3示出了根据本公开的各种实施例的示例机器学习模型;

[0019] 图4示出了根据本公开的各种实施例的示例语义解析过程;

[0020] 图5示出了根据本公开的各种实施例的具有示例话语的示例机器学习模型;

[0021] 图6示出了根据本公开的各种实施例的示例机器学习模型训练过程;

[0022] 图7示出了根据本公开的各种实施例的对机器学习模型的示例IOB格式支持;以及

[0023] 图8A、图8B、图8C、图8D和图8E示出了根据本公开的各种实施例的示例机器学习模型性能图。

## 具体实施方式

[0024] 在描述“具体实施方式”之前,阐明在整个专利文件中使用的某些单词和短语的定义可能是有利的。术语“发送”、“接收”和“通信”及其派生词涵盖直接和间接通信。术语“包括”和“包含”及其派生词是指包括但不限于。术语“或”是包含性的,意味着和/或。短语“与……关联”及其派生词表示包括、包含在其中、与之互连、包含、包含在其中、连接到或与……连接、耦接到或与……耦接、与……通信、与……合作、交织、并置、与……近似、受其约束或具有、具有、具有……的特性、和…有关系或者与…有关系等。

[0025] 此外,下面描述的各种功能可以由一个或更多个计算机程序实现或支持,每个计算机程序由计算机可读程序代码形成并记录在计算机可读介质中。术语“应用程序”和“程序”是指适用于在合适的计算机可读程序代码中实现的一个或更多个计算机程序、软件组件、指令集、过程、函数、对象、类、实例、相关数据或其一部分。短语“计算机可读程序代码”

包括任何类型的计算机代码,包括源代码、目标代码和可执行代码。短语“计算机可读介质”包括能够由计算机访问的任何类型的介质,例如只读存储器 (ROM)、随机存取存储器 (RAM)、硬盘驱动器、光盘 (CD)、数字视频光盘 (DVD) 或任何其他类型的存储器。“非暂时性”计算机可读介质不包括传输瞬时电信号或其他信号的有线、无线、光学或其他通信链路。非暂时性计算机可读介质包括可永久存储数据的介质以及可存储数据并随后覆盖的介质,例如可重写光盘或可擦除存储设备。

[0026] 如本文所使用的,诸如“具有”、“可能具有”、“包括”或“可能包括”特征(例如数字、功能、操作或诸如零件之类的组件)之类的术语和短语表示该特征的存在,并且不排除其他特征的存在。此外,如本文所使用的,短语“A或B”、“A和/或B中的至少一个”或“A和/或B中的一个或多个”可以包括A和B的所有可能的组合。例如,“A或B”、“A和B中的至少一个”和“A或B中的至少一个”可以表示以下所有情况:(1) 包括至少一个A,(2) 包括至少一个B,或者(3) 包括至少一个A和至少一个B。此外,如本文中所使用的,术语“第一”和“第二”可以修改各种组件而与重要性无关,并且不限制这些组件。这些词语仅用于区分一个组件与另一个组件。例如,第一用户设备和第二用户设备可以表示彼此不同的用户设备,而不管这些用户设备的顺序或重要性。在不脱离本公开的范围的情况下,第一组件可以被称为第二组件,反之亦然。

[0027] 应当理解,当元件(例如第一元件)被称为“与另一元件(例如,第二元件)操作地或通信地耦接/连接”或“连接到另一元件(例如,第二元件)时”,该元件可以直接地耦接/连接到或通过第三元件耦接/连接到该另一元件。相反,应当理解,当元件(例如第一元件)被称为“直接耦合到”或“直接连接到”另一元件(例如第二元件)时,该元件与该另一元件之间没有其它元件(例如第三元件)介入。

[0028] 如本文所使用的,短语“配置(或设置为)”可以根据情况与短语“适合”、“具有……能力”、“设计为”、“适应于”、“制造为”或“能够”交换使用。短语“配置为”并不实质上表示“被以硬件方式具体设计用于”。相反,短语“配置为”可以指示设备能够与其他设备或组件一起执行特定操作。例如,短语“被配置(或被设置)为执行A、B和C的处理器”可以表示可以通过执行存储在存储器设备中的一个或多个软件程序来执行操作的通用处理器(诸如,CPU或应用处理器),或用于执行操作的专用处理器(诸如,嵌入式处理器)。

[0029] 本文使用的术语和短语被提供仅用于描述本公开的一些实施例,而不是限制本公开的其它实施例的范围。应理解,除非情景中另有明确指示,单数形式的“一”、“该”和“所述”也包括多个所指对象。本文所使用的所有术语和短语,包括技术和科学术语和短语,具有与本公开的实施例所属领域的普通技术人员通常理解的含义相同的含义。还将进一步理解,诸如通常使用的词典中定义的术语和短语的术语和短语应该被解释为具有与它们在相关领域的上下文中的含义一致的含义,并且将不会以理想化的或者过于正式的意义来解释,除非本文明确地另外定义。在一些情况下,本文中定义的术语和短语可以被解释为排除本公开的实施例。

[0030] 根据本公开的实施例的“电子设备”的示例可以包括以下中的至少一项:智能电话机、平板个人计算机(PC)、移动电话机、视频电话机、电子书阅读器、台式PC、膝上型计算机、上网本计算机、工作站、个人数字助理(PDA)、便携式多媒体播放器(PMP)、MP3播放器、移动医疗设备、相机或可穿戴设备(诸如智能眼镜、头戴式设备(HMD)、电子服装、电子手镯、电子

项链、电子配件、电子纹身、智能镜子、或智能手表)。电子设备的其它示例包括智能家用电器。智能家居电器的示例可以包括以下中的至少一项:电视机、数字视频盘(DVD)播放器、音频播放器、冰箱、空调、吸尘器、烤箱、微波炉、洗衣机、烘干机、空气净化器、机顶盒、家庭自动化控制面板、安全控制面板、电视盒(诸如,SAMSUNG HOMESYNC、APPLE TV、或者GOOGLE TV)、游戏控制台(诸如,XBOX、PLAYSTATION或NINTENDO)、电子词典、电子钥匙、摄录像机或电子相框。电子设备的其它示例包括以下中的至少一项:各种医疗设备(诸如,各种便携式医疗测量设备(如血糖测量设备、心跳测量设备或体温测量设备)、磁共振血管造影术(MRA)设备、磁共振成像(MRI)设备、计算机断层摄影(CT)设备、成像设备或超声设备)、导航设备、全球定位系统(GPS)接收器、事件数据记录器(EDR)、飞行数据记录器(FDR)、汽车信息娱乐设备、航海电子设备(诸如,航海导航设备或陀螺罗盘)、航空电子设备、安全设备、车辆头部单元、工业或家庭机器人、自动取款机(ATM)、销售点(POS)设备或物联网(IoT)设备(诸如,灯泡、各种传感器、电表或燃气表、洒水器、火警报警器、恒温器、路灯、烤面包机、健身设备、热水箱、加热器或锅炉)。电子设备的其它示例包括以下中的至少一部分:一件家具或建筑物/结构、电子板、电子签名接收设备、投影仪或各种测量设备(诸如,用于测量水、电、气体或电磁波的设备)。注意,根据本公开的各种实施例,电子设备可以是上面列出的设备中的一种或组合。根据本公开的一些实施例,电子设备可以是柔性电子设备。本文所公开的电子设备不限于上述设备,并且取决于技术的发展可以包括新的电子设备。

[0031] 在以下描述中,根据本公开的各种实施例,参考附图描述电子设备。如本文所使用的,术语“用户”可以表示使用该电子设备的人或另一设备(诸如,人工智能电子设备)。

[0032] 贯穿本专利文献可以提供对于其它某些词语和短语的定义。本领域普通技术人员将理解,在许多实例中,即使不是在大多数实例中,这样的定义适用于这样定义的词语和短语的先前的使用以及将来的使用。

[0033] 本申请中的任何描述都不应被理解为暗示任何特定的元件、步骤或功能是必须包括在权利要求范围内的必要元件。专利主题的范围仅由权利要求限定。在权利要求之内的任何其它术语的使用,包括但不限于“机制”、“模块”、“设备”、“单元”、“组件”、“元件”、“构件”、“装置”、“机器”、“系统”、“处理器”或“控制器”,应当被申请人理解为指代相关技术领域的技术人员已知的结构。

[0034] 下面讨论的图1到图8E以及本公开的各种实施例是参考附图描述的。然而,应当理解,本公开不限于这些实施例,对其的所有变化和/或等效物或替代物也属于本公开的范围。

[0035] 本公开的各种实施例通过利用普遍存在的用户信息来减轻现有语义框架解析模型的缺点。在一些实施例中,基于渐进式注意力的循环神经网络(RNN)模型或其他机器学习系统注释信息类型并提取每种信息类型的相关先验知识以继续学习意图检测和槽位填充。在特定实施例中,主RNN结构与一组不同的层堆叠在一起,该一组不同的层以渐进的方式被逐一训练。可以利用一个或更多个电子设备(例如本公开中描述的电子设备)来建立基于渐进式注意力的RNN。

[0036] 语义框架解析可以由使用RNN的这些相同电子设备或由不同的电子设备来操作。通过利用用户信息,本公开的各种实施例显著提高了性能并且显著减少了使用带注释的训练集的需要。此外,本公开的各种实施例缩短了用于实现一定性能(例如达到与现有语义框

架解析模型相同的性能)的训练时间,同时使用明显更少的带注释的训练样本(例如,带注释的训练样本减少了80%)。获得类似性能的训练时间也可以显著减少(例如减少60%以上)。因此,本公开使得能够在新域中以较少的带注释的训练集来更快地开发语义框架解析模型。

[0037] 图1示出了根据本公开的各种实施例的示例网络配置100。图1所示的网络配置100的实施例仅用于说明。在不脱离本公开的范围的情况下,可以使用网络配置100的其他实施例。

[0038] 根据本公开的实施例,电子设备101被包括在网络环境100中。电子设备101可以包括总线110、处理器120、存储器130、输入/输出(I/O)接口150、显示器160、通信接口170或事件处理模块180中的至少一项。在某些实施例中,电子设备101可以排除至少一个组件,或者可以添加另一组件。

[0039] 总线110包括用于将上述组件120至180彼此连接并且在上述组件之间传送通信(例如,控制消息和/或数据)的电路。处理器120包括中央处理器(CPU)、应用处理器(AP)和通信处理器(CP)中的一个或多个。处理器120能够对电子设备101的至少一个其他组件执行控制和/或执行与通信有关的数据处理或操作。在一些实施例中,处理器可以是图形处理器单元(GPU)。根据本公开的各种实施例,处理器120可以例如通过接收一个或多个语言情景、将话语的至少一部分分类为信息类型、以及基于一个或多个语言情景和话语的已分类的至少一部分接收并填充一个或多个槽位,来训练或执行机器学习模型。

[0040] 存储器130可以包括易失性存储器和/或非易失性存储器。例如,存储器130可以存储于电子设备101的至少一个其他组件有关的命令或数据。根据本公开的实施例,存储器130可以存储软件和/或程序140。程序140包括例如内核141、中间件143、应用编程接口(API) 145和/或应用程序(或“应用”) 147。内核141、中间件143和API 145中的至少一部分可以被称为操作系统(OS)。

[0041] 内核141可以控制或管理用于执行在其他程序(例如,中间件143、API 145或应用程序147)中实现的操作或功能的系统资源(例如,总线110、处理器120、存储器130等)。内核141提供用于允许中间件143、API 145或应用147访问电子设备101的各个单独组件以控制或管理系统资源的接口。应用147可以包括接收输入话语数据以提供包括语音识别和建议服务或其他服务的服务(诸如虚拟助手服务)的应用。应用147还可以包括机器学习应用,其用于对话语数据进行分类、使用分类的话语数据填充槽位以及通过本公开中描述的各种过程从话语数据确定意图。

[0042] 中间件143可以用作例如中继器,以允许API 145或应用147与内核141通信数据。可以提供多个应用147。中间件143能够例如通过将使用电子设备101(诸如总线110、处理器120或存储器130)的系统资源的优先级分配给多个应用147中的至少一个,来控制从应用147接收的工作请求。

[0043] API 145是允许应用147控制从内核141或中间件143提供的功能的接口。例如,API 145包括用于归档控制、窗口控制、图像处理或文本控制的至少一个接口或功能(例如命令)。

[0044] I/O接口150用作例如可以将从用户或其他外部设备输入的命令或数据输入到电子设备101的其他组件的接口。此外,I/O接口150可以向用户或其他外部设备输出从电子设备



101的其他组件接收的命令或数据。

[0045] 显示器160可以包括例如液晶显示器 (LCD)、发光二极管 (LED) 显示器、有机发光二极管 (OLED) 显示器、量子点发光二极管 (QLED) 显示器、微机电系统 (MEMS) 显示器或电子纸显示器。显示器160也可以是深度感知显示器,例如多焦点显示器。显示器160可以向用户显示各种内容 (诸如文本、图像、视频、图标或符号)。显示器160可以包括触摸屏,并且可以使用电子笔或用户的身体部位来接收例如触摸、手势、接近或悬停输入。

[0046] 通信接口170例如能够在电子设备101与外部电子设备 (诸如第一电子设备102、第二电子设备104或服务器106) 之间建立通信。例如,通信接口170可以通过无线或有线通信与网络162或网络164连接以与外部电子设备通信。通信接口170可以是有线或无线收发器或用于发送和接收信号 (例如由通信接口170接收的与话语数据信号、信息类型分类信号或其他信号有关的信号) 的任何其他组件。

[0047] 电子设备101进一步包括一个或更多个传感器,其能够计量物理量或检测电子设备101的激活状态并将计量或检测到的信息转换为电信号。例如,传感器可以包括一个或更多个用于触摸输入的按钮、一个或更多个相机、手势传感器、陀螺仪或陀螺仪传感器、气压传感器、磁传感器或磁力计、加速度传感器或加速计、握持传感器、接近传感器、颜色传感器 (例如红绿蓝 (RGB) 传感器)、生物物理传感器、温度传感器、湿度传感器、照度传感器、紫外线 (UV) 传感器、肌电图 (EMG) 传感器、脑电图 (EEG) 传感器、心电图 (ECG) 传感器、IR传感器、超声传感器、虹膜传感器、指纹传感器等。传感器2可以包括用于控制所包括的至少一个传感器的控制电路。这些传感器中的任何一个都可以位于电子设备101内。

[0048] 该无线通信能够使用例如作为蜂窝通信协议的长期演进 (LTE)、高级长期演进 (LTE-A)、第五代无线系统 (5G)、毫米波或60GHz无线通信、无线USB、码分多址 (CDMA)、宽带码分多址 (WCDMA)、通用移动通信系统 (UMTS)、无线宽带 (WiBro) 或全球移动通信系统 (GSM) 中的至少一种。有线通信可以包括以下各项中的至少一项:通用串行总线 (USB)、高清多媒体接口 (HDMI)、推荐标准232 (RS-232) 或普通老式电话业务 (POTS)。网络162可以包括至少一个通信网络,例如计算机网络 (例如局域网 (LAN) 或广域网 (WAN))、互联网或电话网络。

[0049] 第一外部电子设备102和第二外部电子设备104以及服务器106均可以是与电子设备101的类型相同或不同的设备。根据本公开的某些实施例,服务器106包括一组一个或更多个服务器。而且,根据本公开的某些实施例,在电子设备101上执行的所有或一些操作可以在另一个或更多个其他电子设备 (诸如电子设备102和电子设备104或服务器106) 上执行。此外,根据本公开的某些实施例,当电子设备101自动地或者响应于请求而应该执行特定功能或服务时,代替其自身执行该功能或服务或者在其自身执行该功能或服务之外,电子设备101可以请求向其他设备 (例如,电子设备102和电子设备104或服务器106) 执行该功能相关的至少一部分功能。该其他电子设备 (例如,电子设备102和电子设备104或服务器106) 能够执行所请求的功能或附加功能,并可以向电子设备101传送执行的结果。电子设备101可以自己使用接收到的结果,或者附加地处理接收到的结果,以提供所请求的功能或服务。为此,可使用例如云计算技术、分布式计算技术或客户机-服务器计算技术。虽然图1示出了电子设备101包括通信接口170以经由网络162与外部电子设备104或服务器106通信,但是根据本公开的实施例,电子设备101可以独立地操作而没有单独的通信功能。

[0050] 服务器106可以包括与电子设备101相同或相似的组件110至180 (或其合适的子

集)。服务器106可以通过执行在电子设备101上实现的操作(或功能)中的至少一项来支持驱动电子设备101。例如,服务器106可以包括可以支持在电子设备101中实现的处理器120的处理模块或处理器。服务器106还可以包括事件处理模块(未示出),该事件处理模块可以支持在电子设备101中实现的事件处理模块180。例如,事件处理模块180可以处理从其他元件(例如,处理器120、存储器130、输入/输出接口150或通信接口170)获得的信息的至少一部分,并且可以将其以各种方式提供给用户。

[0051] 尽管在图1中事件处理模块180被示为与处理器120分离的模块,但是事件处理模块180的至少一部分可以被包括或实现在处理器120或至少一个其他模块中,或者事件处理模块180的整体功能可以被包括或实现在处理器120或另一处理器中。事件处理模块180可以与存储在存储器130中的至少一个程序140互操作地执行根据本公开的实施例的操作。

[0052] 尽管图1示出了网络配置100的一个示例,但可以对图1进行各种更改。例如,网络配置100可以以任何合适的布置包括任意数量的每个组件。通常,计算和通信系统具有各种各样的配置,并且图1不将本公开的范围限制为任何特定的配置。尽管图1示出了可以使用本专利文件中公开的各种特征的一个操作环境,但是这些特征可以用在任何其他合适的系统中。

[0053] 图2示出了根据本公开的各种实施例的示例机器学习架构200。如图2所示,架构200包括基于渐进式注意力的模型202。模型202被配置为接收话语204并确定和输出意图类别206和槽位标记208。模型202还被配置为从各种源检索信息以改善性能。在该示例中,这些源包括当前用户信息210,例如用户偏好212和当前情景214。

[0054] 当前用户信息210提供与向模型202提供话语204的用户有关的个性化信息。现有的个人助理旨在针对任何用户和在任何情况下对相同的话语执行相同的意图。这样的通用NLU引擎可能无法满足每个用户的不同需求,并且可能导致在不同情景中出现错误甚至无法实现的操作。作为一个特定示例,三个不同的用户可以说“找到McDonald”,而每个用户的意图都不同。可以通过每个用户的用户偏好212来了解话语204背后的意图。这三个用户中第一用户的意图可能是查找名为“McDonald”的人的联系人信息。该第一用户的用户偏好212可以指示该第一用户具有经常使用或引用的名为“McDonald”的存储联系人,因此该用户偏好212可以使模型202了解到:对于第一用户意图关联该话语和联系人域,至少存在较高的可能性。如果三个用户中的第二用户经常去McDonalds吃饭,则第二用户的用户偏好212可以使模型202了解到:对于第二用户意图关联该话语和餐厅域,至少存在较高的可能性。如果三个用户中的第三用户经常听名为“Audra McDonald”的艺术家的音乐,则第三用户的用户偏好212可以使模型202了解到:对于第三用户意图关联该话语和音乐域,至少存在较高的可能性。

[0055] 架构200还包括外部知识216,其可以将用户或其他数据的当前情景使模型202了解。例如,为了确定用户偏好212,模型202可以例如经由该应用的应用编程接口(API)检索用户所使用的应用的用户偏好数据,以从用户操作的设备或远程设备检索诸如使用频率、用户搜索或输入、或者其他数据。

[0056] 作为另一特定示例,三个不同的用户可以说出“给我查找出行”一词,每个用户的意图不同。可以通过每个用户的当前情景214来了解话语204后面的意图。如果三个用户中的第一用户的位置是旧金山,则当前情景214可以使模型202了解到:对于第一用户意图关

联该话语和(在旧金山地区提供的)第一共享出行服务的域,存在较高的可能性。如果三个用户中的第二用户当前位于法国巴黎,则当前情景214可以使模型202了解到:对于第二用户意图关联该话语和(在巴黎提供的)第二共享出行服务的域,存在较高的可能性。如果三个用户中的第三用户位于中国北京,则当前情景214可以使模型202了解到:对于第二用户意图关联该话语和(在北京提供的)第三共享出行服务的域,存在较高的可能性。外部知识216可以向模型202通知用户的当前情景或其他数据,例如用户的位置。例如,为了确定用户的当前情景214,模型202可以从外部位置或基于全球定位系统(GPS)的服务(例如,GOOGLE MAPS距离矩阵API)中检索用户的位置。

[0057] 架构200还包括信息字典218。模型202使用信息字典218定义用户信息类型与NLU槽位之间的类别关系。在一些实施例中,词典218中的每个关键词是用户信息的类型,并且关键词的对应值是属于该类型的槽位。例如,字典218中的用户信息的类型可以是位置信息类型(例如“@location”),并且与该位置类型相关联的槽位可以包括“从”(from)位置槽位(“@fromloc”)和一个“到”(to)位置槽位(“@to loc”)。每种类型的用户信息也可以具有关联的测量度量。例如,对于位置信息类型,测量度量可以是地理距离度量,使得通过使用外部知识216,当前用户位置可以被确定并被用于确定当前用户位置与说出的位置之间的地理距离以及从信息字典分配的位置类型。

[0058] 作为另一示例,词典218中的用户信息的类型可以是用户应用情景类型。与应用情景类型关联的槽位可以包括诸如联系人槽位、餐厅槽位、音乐槽位、共享出行槽位或其他槽位。每种类型的用户信息也可以具有关联的测量度量。例如,对于应用情景类型,测量度量可以是用户偏好度量,使得通过使用外部知识216,针对不同应用的当前用户偏好可以被确定(例如,基于用户应用使用频率)并被用于针对一个或更多个应用确定频率度量值。架构200因此合并用户偏好、情景信息、外部知识和槽位类别,从而可以在不使用大量历史个人话语数据的情况下确定用户的优选意图。

[0059] 尽管图2示出了机器学习架构200的一个示例,但可以对图2进行各种更改。例如,对于机器学习架构200的每种单独的实施方式,机器学习架构200不一定需要访问图2所示的所有各种类型的数据。机器学习架构200的其他实现方式可以省略这些类型的数据中的一个或更多个,并且机器学习架构200可以使用其他或其他类型的数据。

[0060] 图3示出了根据本公开的各种实施例的示例机器学习模型300。意图检测和槽位填充是由模型300执行的两个任务,用于在NLU应用中进行语义框架解析。语义框架解析的一个示例目标或功能是通过正确标识用户意图并基于意图标记槽位来了解用户话语的变化。可以将意图检测视为话语分类问题,可以使用分类器(例如基于支持向量机(SVM)和递归神经网络(RNN)的模型)对其进行建模。作为序列标记问题,可以使用机器学习方法(例如最大熵马尔可夫模型和条件随机字段(CRF))以及基于RNN的方法(以逐个话语标记每个单词)来解决槽位填充问题。

[0061] 如图3所示,模型300包括话语编码器层302、用户信息标记或信息类型分类层304、槽位填充层306和意图检测层308。给定输入话语作为长度T的序列x,意图检测识别x的意图类别I,并且槽位填充将x映射到相同长度T的相应标记序列y。如图3所示,模型300中每个单元的出现次数可以取决于话语中的话语令牌的数量或话语的长度。

[0062] 在一些实施例中,模型300是一起学习两个任务的联合模型,例如通过包括双向

RNN作为话语编码器层302来学习。在这些实施例中,话语编码器310可以表示一个或多个长短期记忆(LSTM)单元,该单元读取输入话语为序列 $x$ 。在每个时间戳 $t$ 处,学习情景向量 $c_t$ 被学习以与RNN隐藏状态 $h_t$ - $c_t \oplus h_t$ -连接,以针对于此槽位标记 $y_t$ 来学习槽位意图。槽位填充层306的隐藏状态用于预测意图检测层308中的意图标记。在这些类型的实施例中,可以如下定义模型300的训练目标以学习联合分布。

$$[0063] \quad P(y|x) = \max_{\theta_r, \theta_s, \theta_l} \prod_{t=1}^T P(y_t | y_1, \dots, y_{t-1}, x; \theta_r, \theta_s, \theta_l) \quad (1)$$

[0064] 在式(1)中, $\theta_r, \theta_s, \theta_l$ 分别是话语编码器层302、槽位填充层306和意图检测层308的可训练参数。在一些实施例中,信息类型分类层304可以使用信息字典218来确定与每个槽位类型相关联的信息类型。话语编码器310接收并处理话语令牌,并输出隐藏状态 $h_t$ 和语言情景向量 $c_t$ 。在话语编码器310使用双向LSTM(BiLSTM)单元的一些实施例中,诸如前向和后向状态( $fh_t, bh_t$ )的输出可以双向传递,如图3所示。话语编码器310还可以可选地从分类组件312和槽位填充组件314接收针对先前话语令牌的数据,以便将关于话语中的先前令牌的数据提供给每个后续话语编码器310。话语编码器310输出语言情景 $c_t$ 和隐藏状态 $h_t$ 。每个时间步长 $t$ 的隐藏状态 $h_t$ 是前向状态 $fh_t$ 和向后状态 $bh_t$ 的级联,使得 $h_t = fh_t \oplus bh_t$ 。将理解,话语编码器310的配置可以包括输入门、输出门和遗忘门以处理由话语编码器310接收的数据。

[0065] 信息类型分类层304中的分类组件312在时间 $t$ 处从前一话语编码器310接收隐藏状态 $h_{t-1}$ 和语言情景 $c_{t-1}$ 。例如,如图3所示,话语输出 $h_1$ 和 $c_1$ 、以及 $h_1$ 和 $c_1$ 中的第一话语令牌的话语编码器310被传递给第一话语令牌的分组件312。信息类型分类层304中的分类组件312使用用户信息类型标记每个话语令牌。由于标记是基于输入话语的语言情景的,因此在一些实施例中,在每个时间戳 $t$ 处,通过所有隐藏状态的加权总和来使用语言情景向量 $c_t$ 。

$$[0066] \quad \{h_k\} \quad \forall 1 \leq k \leq T, \text{其中每个 } c_t = \sum_{k=1}^T \alpha_{t,k} h_k, \text{并且其中 } \alpha_t = \text{softmax}(e_t),$$

使得 $\alpha_{t,j} = \frac{\exp(e_{t,j})}{\sum_{k=1}^T \exp(e_{t,k})}$ 。在一些实施例中, $e_{t,k} = g(s_{t-1}^u, h_k)$ 可以从前一隐藏状态 $s_{t-1}^u$

(其被定义为 $h_{t-1}$ 和 $c_{t-1}$ 的级联, $s_{t-1}^u = h_{t-1} \oplus c_{t-1}$ )的前馈神经网络 $g$ 获知。在每个时间步骤 $t$ ,信息类型分类层304输出 $P_u(t)$ ,可以定义如下。

$$[0067] \quad P_u(t) = \text{softmax}(W_u s_t^u); \quad \tilde{z}_t = \arg \max_{\theta_u} P_u(t) \quad (2)$$

[0068] 槽位填充层306将用户信息提取到模型中。槽位填充层306与从话语编码器层302输出的信息类型分类层304共享相同的隐藏状态 $h_t$ 和语言情景 $c_t$ 。在一些实施例中,每个用户信息类型与外部或预先训练的模型相关联,以从外部知识216中提取语义上有意义的先验知识,以供槽位填充层306使用。例如,位置的语义可以由其纬度和经度表示,以使两个位置之间的距离反映实际的地理距离。每个输入序列 $x$ 与其对应的用户信息 $U$ 相关联,该用户信息 $U$ 可以表示为一组元组(信息类型,信息内容)。例如,从外部知识216,可以由模型300确定出:对于特定用户,对于信息类型为“用户位置”,信息内容为“布鲁克林,NY”。对于与每种类型的用户信息有关的标记词,模型300对于每个时间戳 $t$ 使用外部知识216得出先验向量

$d_t = \{d_1 \dots d_t^{|U|}\}$ , 其中  $|U|$  是IOB格式的用户信息类型的数量 (“B”代表槽位的起始字, “I”代表槽位的中间字, “O”表示该字不属于任何槽位)。在一些实施例中, 每个元素  $d_t^j$  可以定义如下。

$$[0069] \quad d_t(j) = \text{sigmoid}(\beta(j) \odot \delta_t^j) \quad (3)$$

[0070] 在式 (3) 中,  $\beta(j)$  是可训练的向量,  $\delta_t^j$  是基于分类类型和先验知识的测量度量, 其中  $j$  是分类类型。针对每个信息类型  $j$  在时间戳  $t$  处定义  $\delta_t^j$  的计算。例如, 如果来自外部知识 216 的数据是位置数据, 并且信息类型是位置类型, 则让  $\delta_t^{loc}$  为距离, 可以定义如下。

$$[0071] \quad \delta_{loc} = \text{dist}([\text{说出的位置}], [\text{外部知识的位置}]) \approx x \text{ 英里} \quad (4)$$

[0072] 作为特定示例, 如果在旅行情景中说出位置 “NYC”, 则可以在信息类型分类层 304 中用位置信息类型 (例如 “B-loc”) 对这种话语令牌进行分类。这表明信息类型是某种位置 (例如 “fromloc” 或 “toloc” 类型)。槽位填充层 306 从信息类型分类层 304 接收该 B 位置信息类型, 以及从话语编码器层 302 接收隐藏状态和语言情景。可以从外部知识 216 确定用户的位置, 例如通过使用 GPS 应用的 API 确定用户的位置。在该特定示例中, 可以将用户的位置确定为 “布鲁克林, NY”。因此可以如下确定测量度量。

$$[0073] \quad \delta_{loc} = \text{dist}(\text{“NYC”}, \text{“布鲁克林, NY”}) \approx 4.8 \text{ 英里}$$

[0074] 可以由模型 300 确定用户说出的 “NYC” 是 “fromloc” 类型还是 “toloc” 类型。在该示例中, 由于使用外部知识 216 确定的用户的情景位置 “布鲁克林, NY” 距离说出的 “NYC” 位置约 4.8 英里, 因此鼓励槽位填充层 306 将位置 “NYC” 分配给 “B-fromloc” 类型。因此, 模型 300 可以将话语令牌分配给适当的槽位, 而无需更多的训练数据来确定用户说出的位置是 “fromloc” 还是 “toloc”。

[0075] 为了将向量  $d_t$  传递到槽位填充层 306 中, 在一些实施例中, 槽位填充层 306 在从信息类型分类层 304 的 softmax 概率分布  $P_u(t)$  上对每个元素  $d_t^j$  和语言情景  $c_t$  进行加权。这确定出话语中用户信息或语言情景类型的重要性, 这有助于预测话语中每个单词的槽位标记。在一些情况下, 在槽位填充层 306 中的每个时间步骤  $t$  处的 LSTM 单元的输入  $\Phi_t$  可以被定义如下。

$$[0076] \quad \Phi_t = P_u^1(t)d_t^1 \oplus \dots \oplus P_u^m(t)d_t^m \oplus P_o(t)c_t \quad (5)$$

[0077] 在式 (5) 中,  $P_u^j(t)$  和  $P_o(t)$  代表第  $t$  个词语被预测为用户信息的第  $j$  种类型和 “O” 的概率, 其中 “O” 不指定任何类型。时间步骤  $t$  的状态  $s_t$  计算为  $h_t \oplus \Phi_t$ , 并且可以如下预测槽位类型或信息类型分类。

$$[0078] \quad P_s(t) = W_s s_t^s; \quad \hat{y}_t = \underset{\theta_s}{\text{argmax}} P_s(t) \quad (6)$$

[0079] 在一些实施例中, 一旦槽位填充层 306 确定了话语令牌的槽位类型, 就将来自槽位填充层 306 的槽位类型和参数传递给意图检测层 308。意图检测层 308 确定并输出话语的意图类别。例如, 如果话语与预订航班有关, 则意图检测层 308 输出 “航班” 意图类别。意图检测

层308还可以通过使用来自槽位填充层306的隐藏状态的级联来生成意图类别标记的概率分布 $P_I$ ,使得 $s^I = s^s \oplus \dots \oplus s_T^s$ 。这可以定义如下。

$$[0080] \quad P_I(x) = P_I(s_I) = \text{softmax}(W_I S^I); \quad \tilde{I} = \underset{\theta_I}{\operatorname{argmax}} P_I(x) \quad (7)$$

[0081] 在一些实施例中,对于信息类型分类层304, $h_t$ 和 $c_t$ 用于标记属于一种信息类型的词语。可以通过利用提取的先验知识而不是再次使用 $h_t$ 和 $c_t$ 来在槽位填充层306中标记这些词语的语义槽位。然后,槽位填充根据 $h_t$ 和 $c_t$ 来标记不属于任何信息类型的其余词语。

[0082] 尽管图3示出了机器学习模型300的一个示例,但可以对图3进行各种更改。例如,机器学习模型300可以包括任何合适数量的话语编码器310、分类组件312和槽位填充组件314。此外,在处理每个话语时,不一定需要机器学习模型300使用外部知识216和信息字典218。

[0083] 图4示出了根据本公开的各种实施例的示例语义解析过程400。为了便于说明,过程400可以描述为由图1中的电子设备101、102、104中的任何一个或更多个处理器120或服务器106执行的或以其他方式使用的。然而,过程400可以由任何合适的设备和任何合适的系统使用。

[0084] 在框402,处理器接收包括至少一个话语令牌的话语。在框404,处理器在诸如话语编码器层302的话语编码器层中编码每个话语令牌。话语编码器310可以处于各种配置中,诸如使用隐藏节点或单元来处理接收到的话语输入。在一些实施例中,话语编码器310使用BiLSTM单元。在一些实施例中,话语编码器310的输出(例如前向和后向状态( $fh_t, bh_t$ ))可以在每个时间步骤 $t$ 双向传递。话语编码器单元还可从分类层单元和槽位填充层单元接收用于另一话语令牌的数据,以便将关于话语中的其他令牌的数据提供给话语编码器单元。在一些实施例中,话语编码器单元输出语言情景 $c_t$ 和隐藏状态 $h_t$ 。每个时间步长 $t$ 的隐藏状态 $h_t$ 可以是前向状态 $fh_t$ 和向后状态 $bh_t$ 的级联,使得 $h_t = fh_t \oplus bh_t$ 。

[0085] 在框406,处理器将来自话语编码器层的输出输入到分类层,例如信息类型分类层304。在分类层中的每个分类单元从前一话语编码器单元接收隐藏状态 $h_t$ 和语言情景 $c_t$ 。例如,用于第一话语令牌的话语编码器单元输出 $h_1$ 和 $c_1$ ,并且处理器将 $h_1$ 和 $c_1$ 传递到用于第一话语令牌的分层单元。在一些实施例中,分类层可以使用信息字典218来确定与槽位类型相关联的信息类型。在判定框408,对于在分类层接收的每个话语令牌,处理器确定话语令牌是否对应于信息词典218中的信息类型或类别。如果否,则在框410,处理器将话语令牌分类为不属于槽位类型,并且如果使用IOB格式则将令牌分配为“O”。否则,在框412,处理器将话语令牌分类为信息类型或类别。分类层中的分类层单元用信息类型标记每个话语令牌。由于标记是基于输入话语的语言情景,因此在一些实施例中,如上所述,通过所有隐藏状态的加权总和,在每个时间戳 $t$ 处使用语言情景向量 $c_t$ ,这可能涉及使用前馈神经网络 $g$ 。

[0086] 在框414,处理器将话语编码器层的输出和分类层的输出输入到槽位填充层(例如槽位填充层306)中。槽位填充层306与从话语编码器层302输出的信息类型分类层304共享相同的隐藏状态 $h_t$ 和语言情景 $c_t$ 。在一些实施例中,每个用户信息类型与外部或预先训练的模型相关联,以从外部知识216中提取语义上有意义的先验知识。

[0087] 在判定框416,处理器基于在分类层分配给话语令牌的信息类型来确定外部知识

是否可用。如果是,则在框418,处理器基于外部知识216确定测量度量。例如,位置的语义可以由其纬度和经度表示,以使两个位置之间的距离反映实际的地理距离。每个输入序列 $x$ 可以与其对应的用户信息 $U$ 相关联,该用户信息 $U$ 可以表示为一组元组。例如,从外部知识216,处理器可以知道:对于用户,对于信息类型为“用户位置”,信息内容为“布鲁克林,NY”。作为另一个示例,如果话语令牌被分类层304分类为名称类型,则处理器可以使用外部知识216来确定应用使用频率的测量度量,以确定名称的最可能的用途。作为又一示例,如果分类层304将话语令牌分类为优选离开时段,则相关联的槽位可以是@depart\_time或@arrive\_time,并且测量度量可以是时差度量。对于与每种类型的用户信息有关的标记的词语,处理器可以对于每个时间戳 $t$ 使用外部知识216来得出先验向量 $d_t = \{d_1 \dots d_t^{|U|}\}$ ,其中 $|U|$ 是IOB格式的用户信息类型的数量。每个元素 $d_t^j$ 可以由上面的式(3)定义。

[0088] 在框420,处理器在槽位填充层对测量度量和来自话语编码器层的输出进行加权。为了将向量 $d_t$ 传递到槽位填充层,处理器在从分类层的softmax概率分布 $P_u(t)$ 上对每个元素 $d_t^j$ 和语言情景 $c_t$ 进行加权。这确定出话语中用户信息或语言情景类型的重要性,这有助于预测话语中每个单词的槽位标记。槽位填充层中每个时间步长 $t$ 的单元的输入可以由上面的式(5)定义,并且在时间步长 $t$ 的状态 $s^s$ 可以计算为 $h_t \oplus \Phi_t$ 并且可以如上面的式(6)中所定义的那样来预测槽位标记或信息类型分类。

[0089] 如果在判定框416,处理器确定没有外部知识可用于由分类层分配的信息类型(例如,如果分类层为话语令牌分配了“0”),则在框426,处理器对话语编码器层的输出进行加权。在一些实施例中,如果话语令牌和信息类型是不同的类型,则处理器可以设置 $\delta_t^j$ 等于-1,以使其对应的 $d_t^j$ 接近于零。

[0090] 在框422,处理器基于在框420或426处加权的输入,使用槽位填充层来填充槽位。在一些实施例中,一旦槽位填充层306确定了话语令牌的槽位类型,则在框424,处理器将槽位类型和参数从槽位填充层306传递到意图检测或输出层(例如意图检测层308)。意图检测层确定并输出话语的意图类别。意图检测层可以通过使用如上式(7)所定义的来自槽位填充层306的隐藏状态的级联,来生成意图类别标记的概率分布 $P_I$ 。该过程在框428处结束。

[0091] 尽管图4示出了语义解析过程400的一个示例,但可以对图4进行各种更改。例如,尽管被示为一系列步骤,但是图4中的各个步骤可以重叠、并行发生、以不同顺序发生、或发生任何次数。

[0092] 图5示出了根据本公开的各种实施例的具有示例话语的示例机器学习模型500。在一些实施例中,图5中的模型500可以以特定配置来表示图3的模型300,并且将理解,模型500示出了输入到模型中的示例话语。如图5所示,模型500接收“NYC与迈阿密之间往返航班”的话语。图5中的模型500正在处理话语令牌“之间”、“NYC”、“和”与“迈阿密”。但是,应注意,模型500的其他组件可用于处理其余的话语令牌。

[0093] 将每个话语令牌应用于模型500的话语编码器层502中的话语BiLSTM单元510。话语BiLSTM单元对每个话语令牌进行编码。在一些实施例中,话语编码器单元的输出(诸如前向和后向状态( $fh_t, bh_t$ ))可以双向传递,如图5所示。话语BiLSTM单元也可以从分类层504中

的分类层单元512和在槽位填充层506中的槽位填充单元514接收数据,以便将话语中的其他令牌的数据提供给BiLSTM单元。语音BiLSTM单元510输出语言情景 $c_t$ 和隐藏状态 $h_t$ 。每个时间步长 $t$ 的每个隐藏状态 $h_t$ 是前向状态 $fh_t$ 和向后状态 $bh_t$ 的级联,使得 $h_t = fh_t \oplus bh_t$ 。

[0094] 在分类层504中,用于话语令牌的分类单元512从话语编码器层502接收针对该令牌的输出,其中输出包括隐藏状态 $h_t$ 和语言情景 $c_t$ 。例如,如图5所示,用于第一话语令牌的话语BiLSTM单元510输出 $h_1$ 和 $c_1$ ,并且分类单元512接收 $h_1$ 和 $c_1$ 。用于第二话语令牌的话语BiLSTM单元510输出 $h_2$ 和 $c_2$ ,并且分类单元512接收 $h_2$ 和 $c_2$ 。用于第三话语令牌的话语BiLSTM单元510输出 $h_3$ 和 $c_3$ ,并且分类单元512接收 $h_3$ 和 $c_3$ 。用于第四话语令牌的话语BiLSTM单元510输出 $h_4$ 和 $c_4$ ,并且分类单元512接收 $h_4$ 和 $c_4$ 。分类层504确定每个话语令牌是否对应于信息词典218中的信息类型或类别。分类单元512对每个话语令牌进行分类。在图5所示的示例中,第一令牌被分类为“0”(表示它不属于槽位类型),第二令牌被分类为位置类型或“B-loc”类型,第三令牌被分类为“0”,第四令牌被分类为位置类型或“B-loc”类型。

[0095] 对于每个话语 $x$ ,可以使用信息字典218定义用户信息序列 $z$ 。在一些实施例中,训练分类单元512以得出 $z$ 。由于标记是基于输入话语的语言情景,因此在一些实施例中,如上所述,通过所有隐藏状态的加权总和,在每个时间戳 $t$ 处使用语言情景向量 $c_t$ 。在每个时间步骤 $t$ ,分类层504可以输出如上式(2)所定义的 $P_u(t)$ 。

[0096] 该示例示出了分类层504可以最初确定某些令牌的槽位类别,并且槽位填充层506使用该信息来更准确地确定令牌所属的类别中的哪个特定槽位。例如,“NYC”和“迈阿密”都被分类层504分类为“B-loc”类型,指示这两个话语令牌都是位置。然后,槽位填充层506可以确定令牌属于哪个位置类型的槽位,例如哪个令牌属于“B-fromloc”槽位,以及哪个令牌属于“B-to loc”槽位。槽位填充层506中的每个槽位填充单元514接收话语编码器层502、 $h_t$ 和 $c_t$ 的输出以及分类层504的输出。例如,如图5所示,用于第一话语令牌的话语BiLSTM单元510输出 $h_1$ 和 $c_1$ ,并且槽位填充单元514接收 $h_1$ 和 $c_1$ 。用于第二话语令牌的话语BiLSTM单元510输出 $h_2$ 和 $c_2$ ,并且槽位填充单元514接收 $h_2$ 和 $c_2$ 。用于第三话语令牌的话语BiLSTM单元510输出 $h_3$ 和 $c_3$ ,并且槽位填充单元514接收 $h_3$ 和 $c_3$ 。用于第四话语令牌的话语BiLSTM单元510输出 $h_4$ 和 $c_4$ ,并且槽位填充单元514接收 $h_4$ 和 $c_4$ 。

[0097] 槽位填充层506使用从分类层504输出的话语令牌的类别类型,来基于外部知识216确定测量度量516。在此示例中,位置“NYC”和“迈阿密”是在旅行情景中说出的,并在分类层504中被分类为“B-loc”,指示信息类型是某种类型的位置,例如“fromloc”或“to loc”类型。槽位填充层506从分类层504接收该B位置信息类型,以及从话语编码器层502接收隐藏状态和语言情景。可以从外部知识216确定用户的位置,例如通过使用GPS应用的API确定用户的位置。在该示例中,将用户的位置确定为“布鲁克林,NY”。因此可以例如通过使用上面的式(3)和式(4)来确定测量度量516。对于与每种类型的用户信息有关的标记的词语,模型500对于每个时间戳 $t$ 使用外部知识216来得出先验向量 $d_t = \{d_1 \dots d_t^{|U|}\}$ 。每个元素 $d_t^j$ 可以由上面的式(3)定义。

[0098] 槽位填充单元514对测量度量输入 $d_t$ 和来自话语编码器层502的输出进行加权。例如, $d_1$ 被输入到用于第一话语令牌的槽位填充单元514中, $d_2$ 是输入到用于第二话语令牌的槽位填充单元514中, $d_3$ 是输入到用于第三话语令牌的槽位填充单元514中, $d_4$ 是输入到用于



第四话语令牌的槽位填充单元514中。为了将向量 $d_t$ 传递到槽位填充层506,槽位填充单元514在从分类层504的softmax概率分布 $P_u(t)$ 上对每个元素 $d_t^j$ 和语言情景 $c_t$ 进行加权。这确定出话语中用户信息或语言情景类型的重要性,这有助于预测话语中每个单词的槽位标记。槽位填充层506中每个时间步长 $t$ 的每个单元的输入可以由上面的式(5)定义,并且在时间步长 $t$ 的状态 $s^s$ 可以计算为 $h_t \oplus \Phi_t$ 并且可以如上面的式(6)中所定义的那样来预测槽位标记或信息类型分类。

[0099] 在该示例中,可以由模型500确定用户说出的“NYC”是“fromloc”类型还是“toloc”类型。由于使用外部知识216的用户位置被确定为与说出的“NYC”位置相距4.8英里,并且由于情景位置为“布鲁克林,NY”,因此鼓励槽位填充层506将位置“NYC”分配给“B-fromloc”类型。以类似的方式,由于用户不在迈阿密附近,因此鼓励模型500将“迈阿密”分配给“B-toloc”类型。因此,模型500可以将话语令牌分配给适当的槽位,而无需更多的训练数据来确定用户说出的每个位置是“fromloc”槽位类型还是“toloc”槽位类型。

[0100] 如果模型500确定不存在可用于由分类层504分配的信息类型的外部知识216(例如,当分类层504将话语令牌分配为“0”时),则在槽位填充层506处的模型500可以对话语编码层502的输出和 $d_t$ 进行加权,以针对“之间”和“与”令牌来加强对“0”的分配,如图5所示。

在一些实施例中,如果话语令牌和信息类型是不同的类型,则处理器可以设置 $\delta_t^j$ 等于-1,以使其对应的 $d_t^j$ 接近于0。

[0101] 当槽位填充层506为每个话语令牌确定了槽位类型时,来自槽位填充层506中的槽位填充单元514的输出将传递到意图检测层508。意图检测层508确定并输出话语的意图类别。在此示例中,由于话语与预订“NYC”和“迈阿密”之间的航班有关,所以意图检测层508输出“航班”意图类别。为了输出该意图类别,意图检测层508可以通过使用如上式(7)所定义的来自槽位填充层506的隐藏状态的级联,来生成意图类别标记的概率分布 $P_I$ 。在下表1中列出了与话语相关联并且在该示例中由模型500处理的信息。

[0102]

话语 ( $x$ )	往返	旅行	航班	之间	NYC	和	迈阿密
槽位 ( $y$ )	B-round_trip	I-round_trip	O	O	B-FromLoc	O	B-ToLoc
意图 ( $I$ )	航班						
用户信息 ( $U$ )	(“用户位置”: “布鲁克林, NY”)						
用户信息序列 ( $z$ )	O	O	O	O	B-Loc	O	B-Loc

[0103] 表1

[0104] 尽管图5以示例话语示出了机器学习模型500的一个示例,但可以对图5进行各种更改。例如,机器学习模型500可以包括任何合适数量的话语编码器510、分类组件512和槽位填充组件514。如上所述,例如,话语编码器510、分类组件512和槽位填充组件514的数量可以(至少部分地)根据包括输入话语的令牌的数量而变化。同样,图5所示的话语类型仅用

于说明。

[0105] 图6示出了根据本公开的各种实施例的示例机器学习模型训练过程600。为了便于说明,过程600可以描述为由图1中的电子设备101、102、104中的任何一个或更多个处理器120或服务器106执行的或以其他方式使用的。然而,过程600可以由任何合适的设备和任何合适的系统使用。

[0106] 在一些实施例中,可以以逐步的方式进行机器学习模型(例如模型300或模型500)的训练。例如,可以利用中间任务来对分类层进行训练(例如信息类型分类层304),然后可以联合训练槽位填充层和意图检测层(例如,槽位填充层306和意图检测层308)。具体地,对于每个话语 $x$ ,可以使用分类层中的信息字典218来定义用户信息序列 $z$ 。在一些实施例中,训练分类单元以得出 $z$ 。然后,可以将具有用户信息的每种类型的语义含义的先验知识提取到模型中,以继续训练意图检测和槽位标记。先验知识可以包括诸如地理距离、应用使用频率的信息,或者在本公开的各种实施例中描述的其他信息。

[0107] 在框602,处理器提供一个或更多个样本话语以用于训练模型,每个样本话语包括一个或更多个标记的令牌,例如表1所示。话语编码器层(例如话语编码器层302)对样本话语进行编码,并输出结果,例如隐藏状态和语言情景。在话语编码器310使用BiLSTM单元的一些实施例中,话语编码器单元的输出可以双向传递。话语编码器单元还可从分类层单元和槽位填充层单元接收用于另一话语令牌的数据,以便将关于话语中的其他令牌的数据提供给话语编码器单元。在一些实施例中,话语编码器单元输出语言情景 $c_t$ 和隐藏状态 $h_t$ 。

[0108] 在框604,处理器将来自话语编码器层的输出输入到分类层中。在分类层中的分类单元从前一话语编码器单元接收隐藏状态 $h_t$ 和语言情景 $c_t$ 。分类层可以使用信息字典218来确定与每个槽位类型相关联的信息类型。在框606,处理器将每个话语令牌分类为信息类型或类别,或者将令牌分类为不属于槽位类别(例如,给令牌分配“0”)。分类层中的分类单元用信息类型来标记每个话语令牌。由于标记是基于输入话语的语言情景,因此一些实施例中,如上所述,通过所有隐藏状态的加权总和,在每个时间戳 $t$ 处使用语言情景向量 $c_t$ ,这可能涉及使用前馈神经网络 $g$ 。

[0109] 在框608,处理器在这一轮训练中基于分类层的输出来以损失 $L_u$ 对分类层进行训练。损失 $L_u$ 可以定义为如下。

$$[0110] \quad L_u(\theta_r, \theta_u) \triangleq -\frac{1}{n} \sum_{i=1}^{|U|} \sum_{t=1}^n z_t^i \log P_u^i(t) \quad (8)$$

[0111] 在式(8)中,  $|U|$  是IOB格式的用户信息类型或类别的数量。在判定框610,处理器确定分类层是否被充分训练。如果否,则过程600返回框604,以在分类层接收由话语编码器层从样本话语生成的更多输出以进行训练。

[0112] 否则,过程600开始对槽位填充层和意图检测层进行联合训练。在框612,处理器从样本话语中检索话语编码器层的输出和分类层的信息类型输出,并将这些输出输入到槽位填充层中。在一些实施例中,每个用户信息类型与外部或预先训练的模型相关联,以从外部知识216中提取语义上有意义的先验知识。处理器基于在分类层分配给话语令牌的信息类型来确定外部知识是否可用。如果可用,则处理器基于外部知识216确定测量度量。在一些实施例中,为了训练模型,可以将样本测量度量提供给模型。对于与用户信息的每种类型有关的标记的词语,处理器针对每个时间戳 $t$ 得出向量 $d_t = \{d_1, \dots, d_t^{|U|}\}$ 。每个元素 $d_t^j$ 可以

由式 (3) 定义。

[0113] 在框614,处理器在槽位填充层对测量度和来自话语编码器层的输出进行加权。为了将向量 $d_t$ 传递到槽位填充层,处理器在从分类层的softmax概率分布 $P_u(t)$ 上对每个元素 $d_t^i$ 和语言情景 $c_t$ 进行加权。槽位填充层中每个时间步长 $t$ 的每个单元的输入可以由上面的式 (5) 定义,并且在时间步长 $t$ 的状态 $s^s$ 可以计算为 $h_t \oplus \Phi_t$ 并且可以如上面的式 (6) 中所定义的那样来预测槽位标记或信息类型分类。

[0114] 在框616,使用槽位填充层的处理器基于提供给槽位填充层的加权输入来填充一个或更多个槽位。在一些实施例中,一旦槽位填充层确定了话语令牌的槽位类型,则在框618处,处理器将槽位类型和参数从槽位填充层传递到意图检测层。意图检测层基于来自槽位填充层的输出来确定并输出话语的意图类别。意图检测层可以通过使用如上式 (7) 所定义的来自槽位填充层的隐藏状态的级联,来生成意图类别标记的概率分布 $P_I$ 。在框620,处理器以损失 $L_s$ 和 $L_I$ 对槽位填充层和意图检测层进行联合训练。在一些实施例中, $L_s$ 和 $L_I$ 可以如下定义。

$$[0115] \quad L_s(\theta_r, \theta_I, \theta_s, \theta_u) \triangleq -\frac{1}{n} \sum_{i=1}^{|S|} \sum_{t=1}^n y_t^i \log P_s^i(t) \quad (9)$$

$$[0116] \quad L_I(\theta_r, \theta_I, \theta_s, \theta_u) \triangleq -\sum_{i=1}^{|I|} I_x^i \log P_I^i(t) \quad (10)$$

[0117] 在式 (9) 中, $|S|$ 是IOB格式的槽位的数量。在式 (10) 中, $|I|$ 是意图的数量。 $P_i$ 代表概率 $P(X=x_i)$ 。在此, $\theta_r, \theta_I, \theta_s, \theta_u$ 分别是话语编码器层、意图检测层、槽位填充层和分类层中的参数。在一些实施例中,在槽位填充层和意图检测层的联合训练期间,可以对先前训练的参数 $\theta_u$ 进行微调。在判定框622,处理器确定槽位填充层和意图检测层是否被充分训练。如果否,则过程600返回到框612以对具有额外样本输入的槽位填充层和意图检测层进行联合训练。否则,过程600在框624处结束。针对图6描述的渐进训练减少了用于训练的带注释的训练集的数量,缩短了训练时间,并且针对新域使用了较少的带注释的训练样本(与其他自然语言理解(NLU)模型相比)。

[0118] 尽管图6示出了机器学习模型训练过程600的一个示例,但可以对图6进行各种更改。例如,尽管被示为一系列步骤,但是图6中的各个步骤可以重叠、并行发生、以不同顺序发生、或发生任何次数。

[0119] 图7示出了根据本公开的各种实施例的对机器学习模型700的示例IOB格式支持。在一些实施例中,模型700可以代表模型300或模型500的至少一部分。由于上面参考图6所述的渐进式训练过程,模型700自然可以支持IOB格式。图7所示的示例示出了话语“纽约”(New York)被模型700通过分类层704的分类单元712将“纽”(New)和“约”(York)分离为单独的令牌而应用于IOB格式,“B-loc”的IOB标记为“纽”,“I-loc”的IOB标记为“约”。这表示“纽”是起始标记,“约”是中间标记或内部标记。

[0120] 一旦分类,令牌就可以被一起提取出地理距离,作为一个示例,可以用于向槽位填充层706中的槽位填充单元714提供输入 $d_t$ 。由于在分类层704的输出 $P_u(t)$ 中B-Loc和I-Loc被认为是不同的标记,因此这些标记可被槽位填充层706用于推断“B-FromLoc”和“I-FromLoc”。针对有分类层704分配的的信息类型,在第 $t$ 个词语被不正确地标记的情况下,由于来自分类层的输出 $P_u(t)$ 将更权重地对 $h_t$ 和 $c_t$ 加权,隐藏状态 $h_t$ 和语言情景 $c_t$ 可以用于推

断槽位标记。此外,对模型700的各层进行联合训练的第二阶段也使模型700学习使用更多的语言情景来纠正任何未正确标记或未正确分类的信息类型。

[0121] 尽管图7示出了机器学习模型700的IOB格式支持的一个示例,但可以对图7进行各种更改。例如,机器学习模型的其他实施例不需要使用IOB格式。

[0122] 图8A、图8B、图8C、图8D和图8E示出了根据本公开的各种实施例的示例机器学习模型性能图。特别地,图8A示出了曲线图801,该曲线图示出了在训练集中使用情景位置信息数据的结果,该结果是使用意图分类准确性百分比测量的。图8B示出了曲线图803,该曲线图示出了在训练集中使用情景位置信息数据的结果,该结果是使用F1分数测量的。图8C示出了曲线图805,该曲线图示出了在训练集中使用情景位置信息和优选时间段数据的结果,该结果是使用意图分类准确性百分比所测量的。图8D示出了曲线图807,该曲线图示出了在训练集中使用情景位置信息和优选时间段数据的结果,该结果是使用F1分数测量的。图8E示出了曲线图809,该曲线图示出了基于用于训练的时段的数量性能的结果,该性能结果是使用F1分数测量的。

[0123] 如图8A、图8B、图8C、图8D和图8E中所示,曲线图801、曲线图803、曲线图805、曲线图807和曲线图809中的每一个都显示了渐进式训练的双向RNN线802,其代表本公开中描述的机器学习模型,例如模型300、模型500或模型700。曲线图801、曲线图803、曲线图805、曲线图807和曲线图809中的每一个还示出了没有用户信息线804的基于注意力的双向RNN基线以及具有用户信息线806的基于注意力的双向RNN基线。如图8A、图8B、图8C和图8D所示,仅使用情景位置信息或结合优选时间段,代表本公开中描述的模型的线802示出了:利用本公开中描述的模型获得的结果具有更高的意图分类器准确性和更高的F1分数,尤其在具有较小训练集大小的情况下。

[0124] 在特定实施例中,使用本公开中描述的模型在意图检测和槽位填充上的性能提升,对于意图检测可以提升高达1.35%,对于槽位填充可以提升高达1.20%。而且,在特定实施例中,在使用比其他模型使用的训练集小80%的训练集时,本公开中描述的模型可以实现与其他模型相似的性能。使用较少的训练数据来获得与其他模型相似的结果,可以更快地开发NLU模型,以处理从用户接收的各种话语。另外,即使当用户的实际个人数据难以收集时,外部知识和信息类型标记的使用也允许个性化和情景NLU模型。

[0125] 如图8E所示,因为可以使用较少的时段来训练本公开中描述的模型,同时获得比其他模型相同或更高的F1分数(由于本公开中描述的渐进式训练过程),所以也可以使用本公开中描述的模型来减少训练时间。在特定实施例中,与其他模型相比,训练时间可以减少60%以上。此外,在特定实施例中,信息类型分类阶段的训练仅可以花费三个时段来实现超过92%的准确性。在一些情况下,更快的训练时间使本公开中描述的模型能够在诸如移动设备的本地设备上实现。在设备上实现模型可以提高模型中使用的个人用户信息的安全性。

[0126] 尽管图8A、图8B、图8C、图8D和图8E示出了机器学习模型性能曲线图的示例,但是可以对这些曲线图进行各种改变。例如,这些曲线图中显示的性能仅用于说明。该专利文件中描述的方法和技术的其他实施方式可以不同于这些附图中所示的方式来操作。此外,以上提供的特定数值(例如百分比,分数和历元数)仅用于说明,并且不将本公开限制于任何特定的实现方式或性能。

[0127] 尽管已经结合示例性实施例描述了本公开,但是本领域技术人员可能得到各种变化和修改的暗示。本公开旨在包括属于所附权利要求的范围内的这些改变和修改。

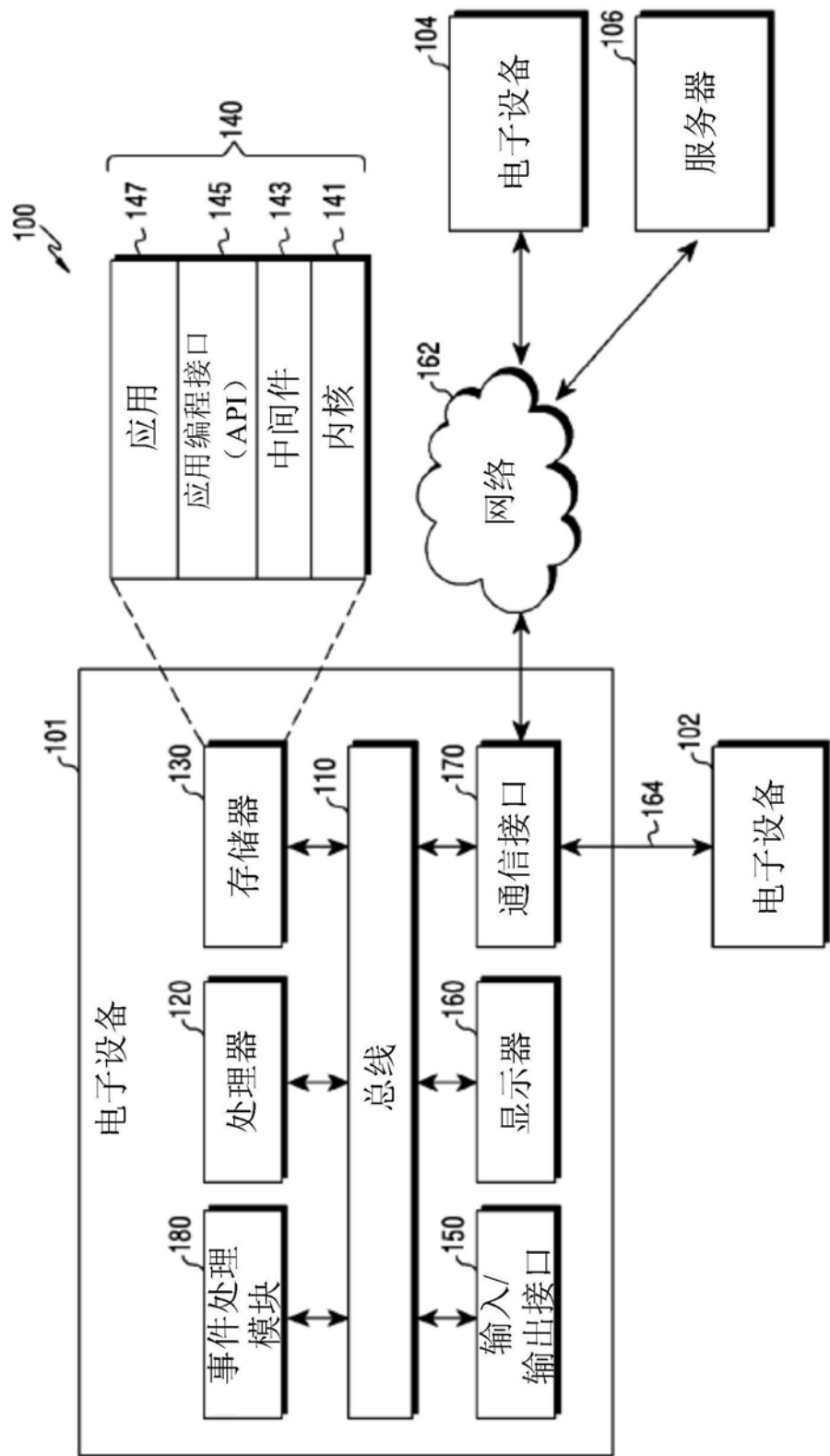


图1

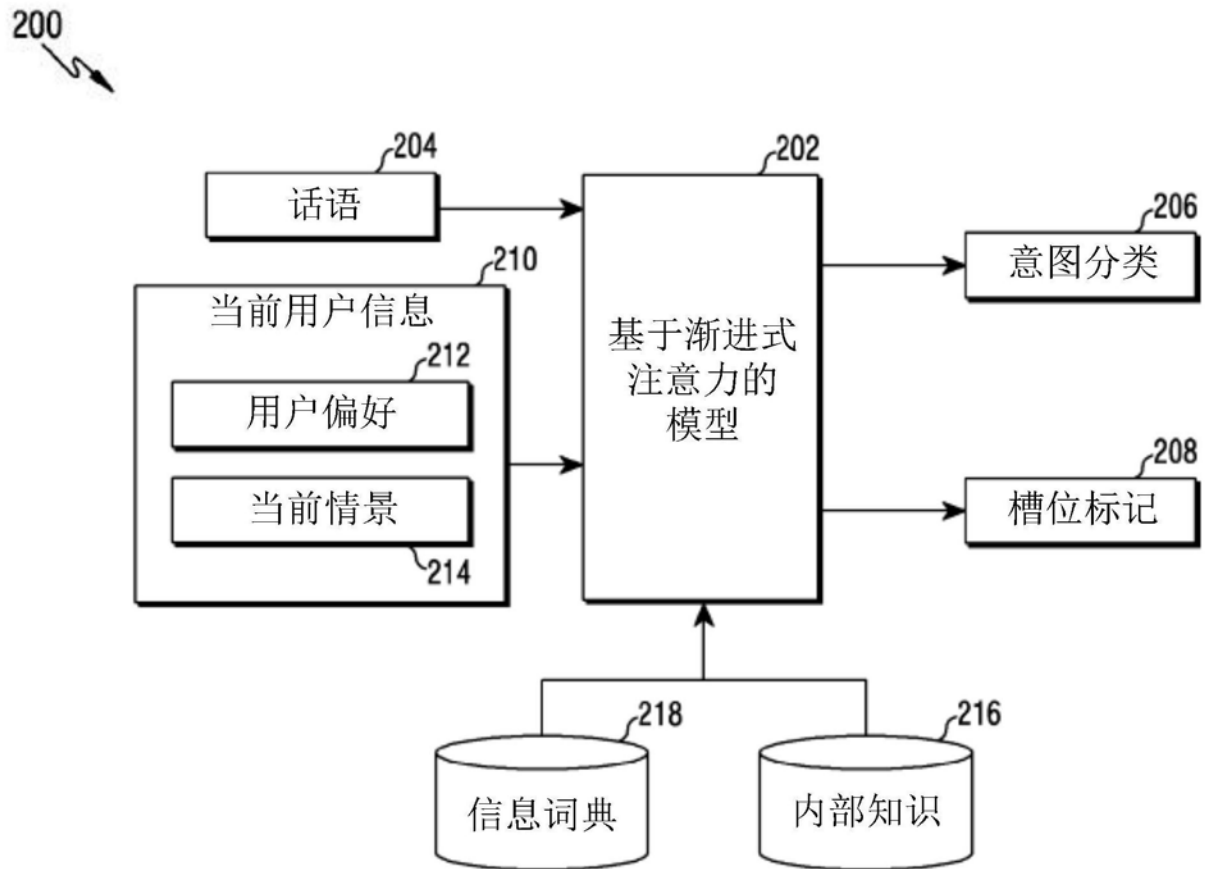


图2

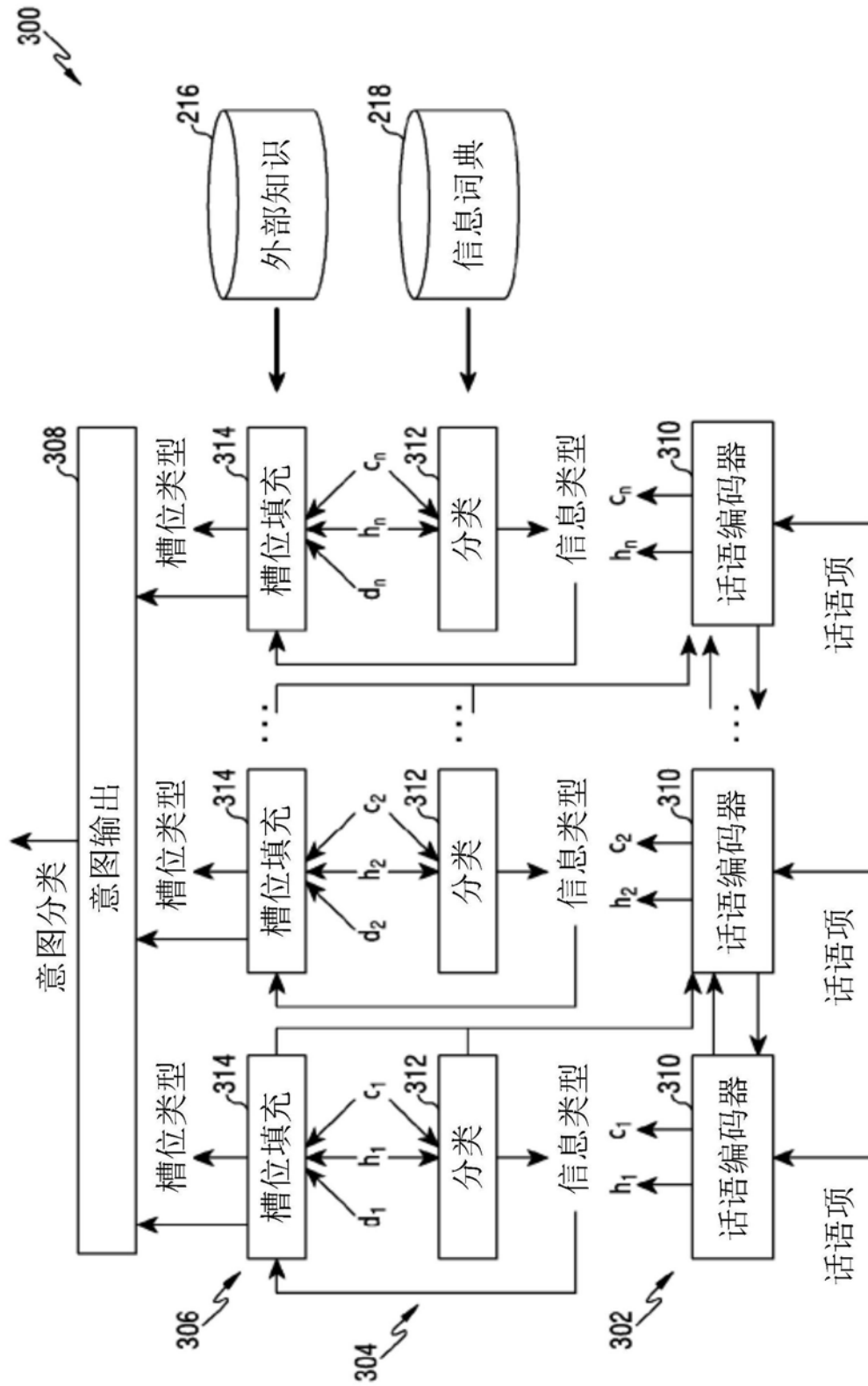


图3



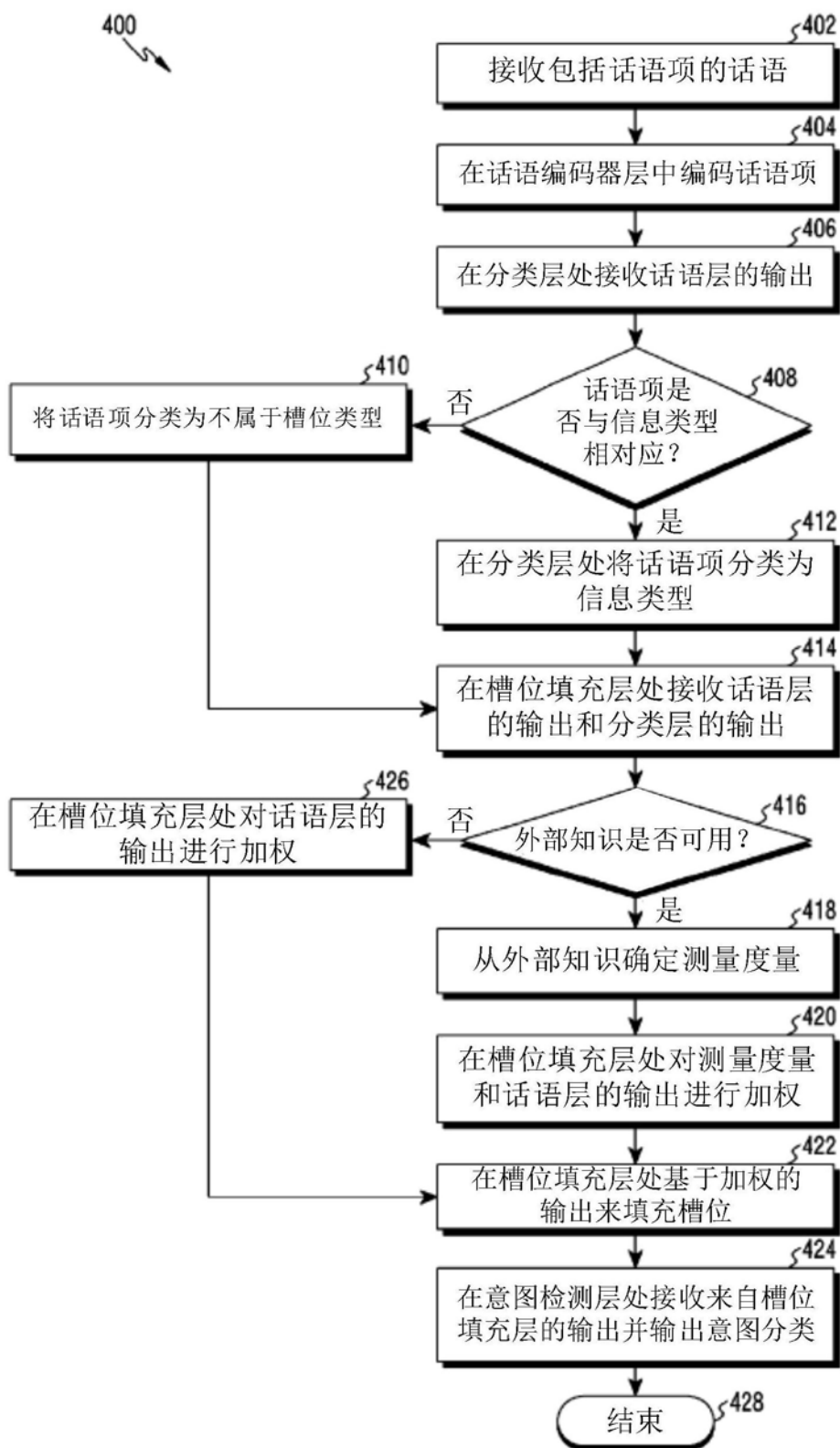


图4

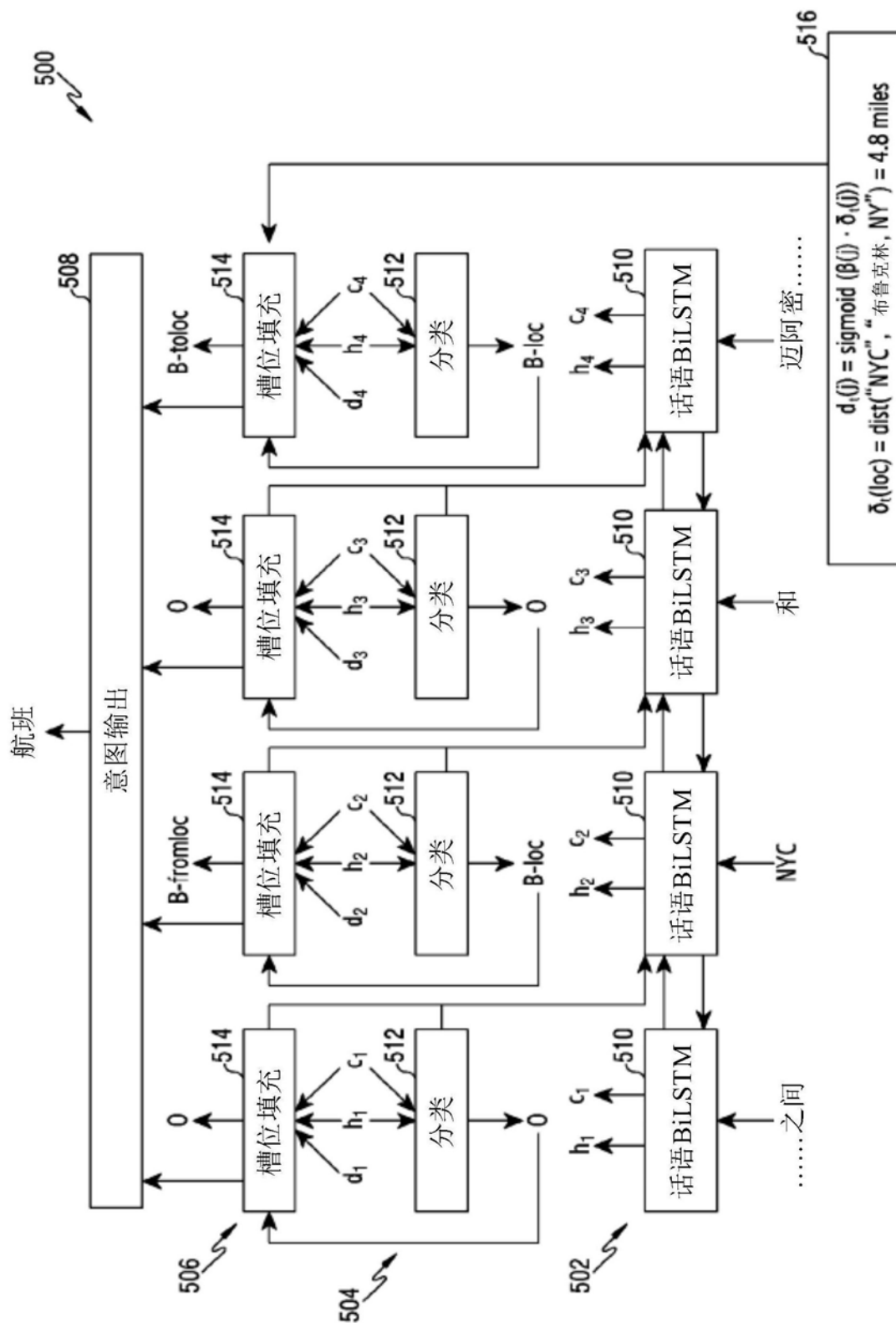


图5

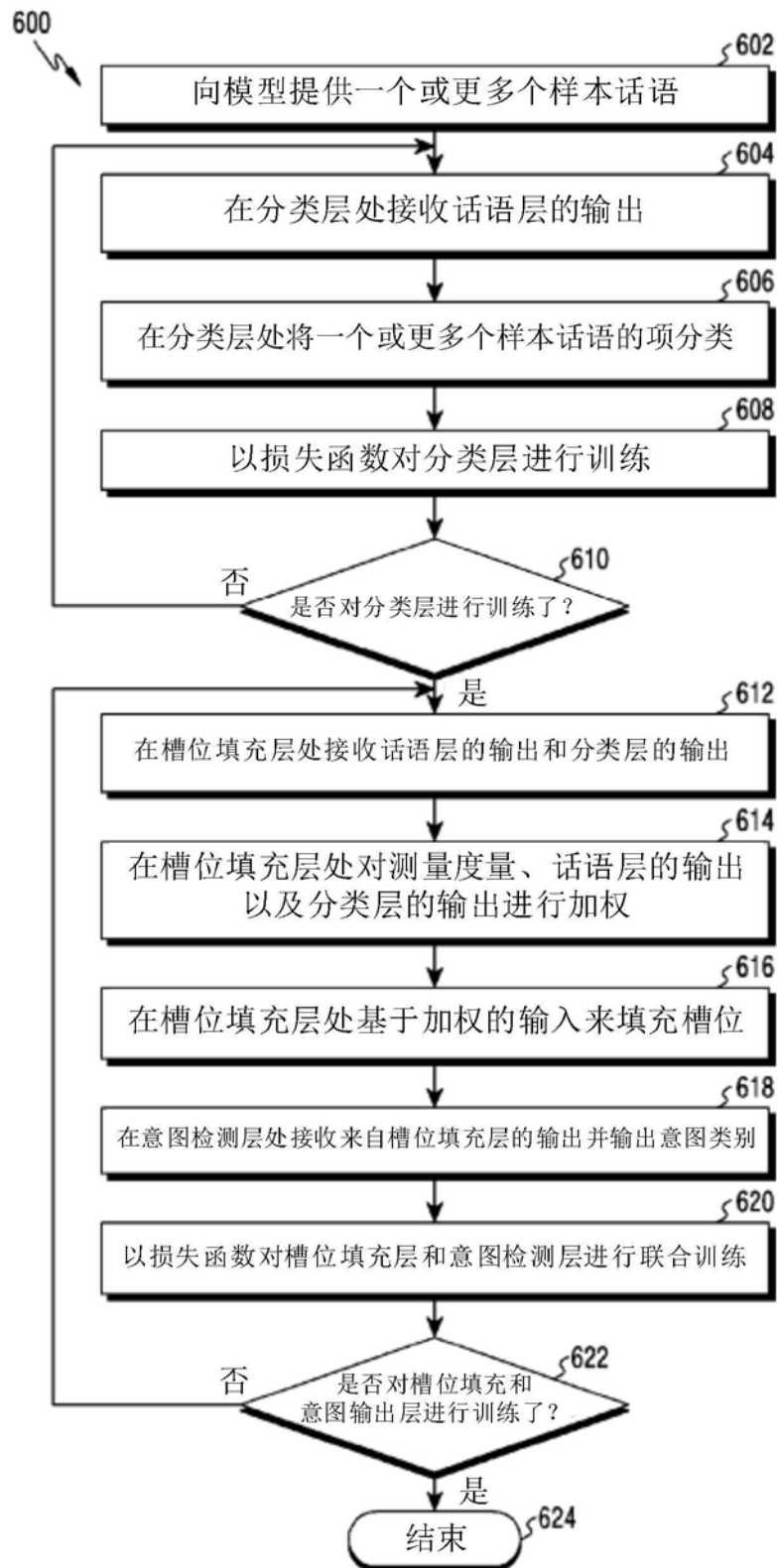


图6

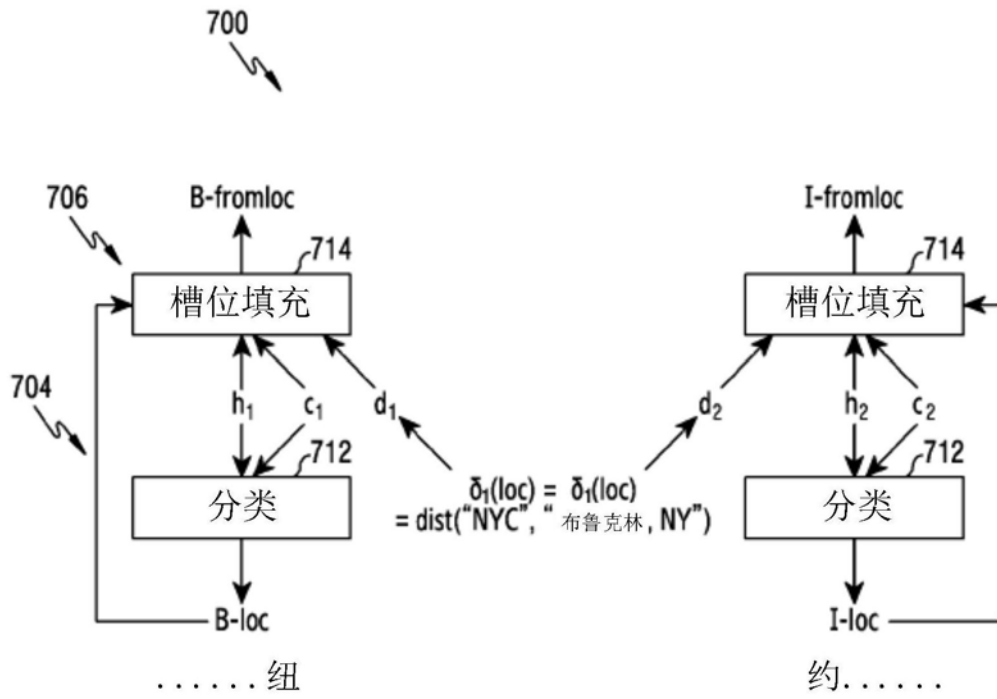


图7

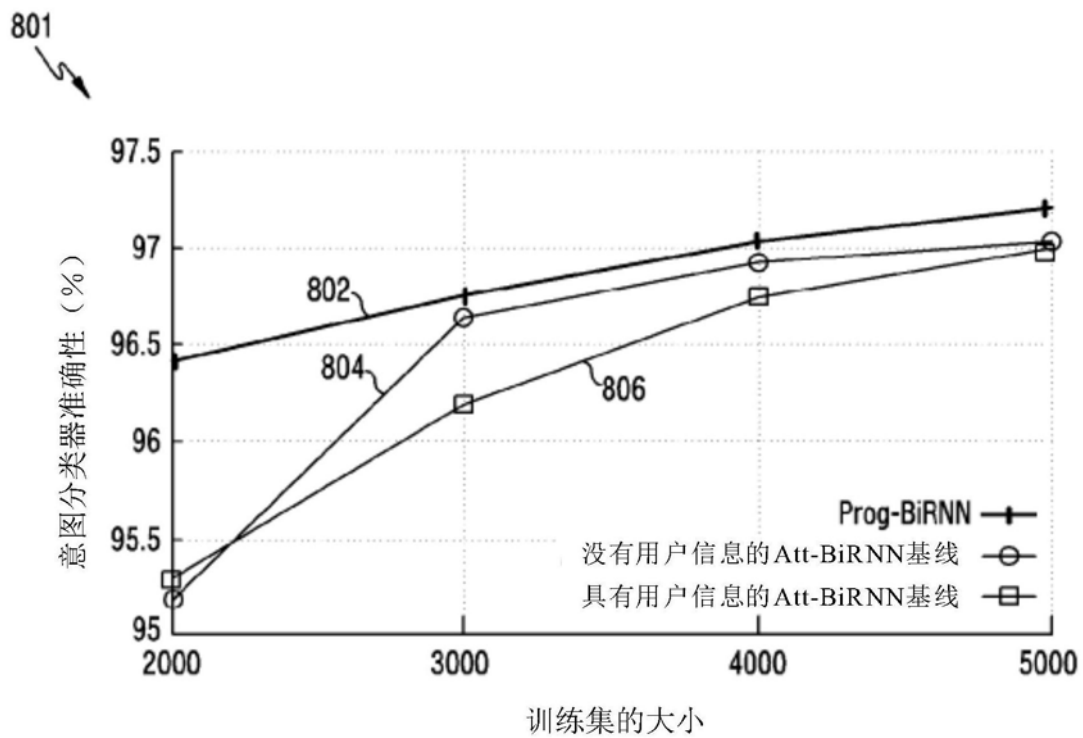


图8A

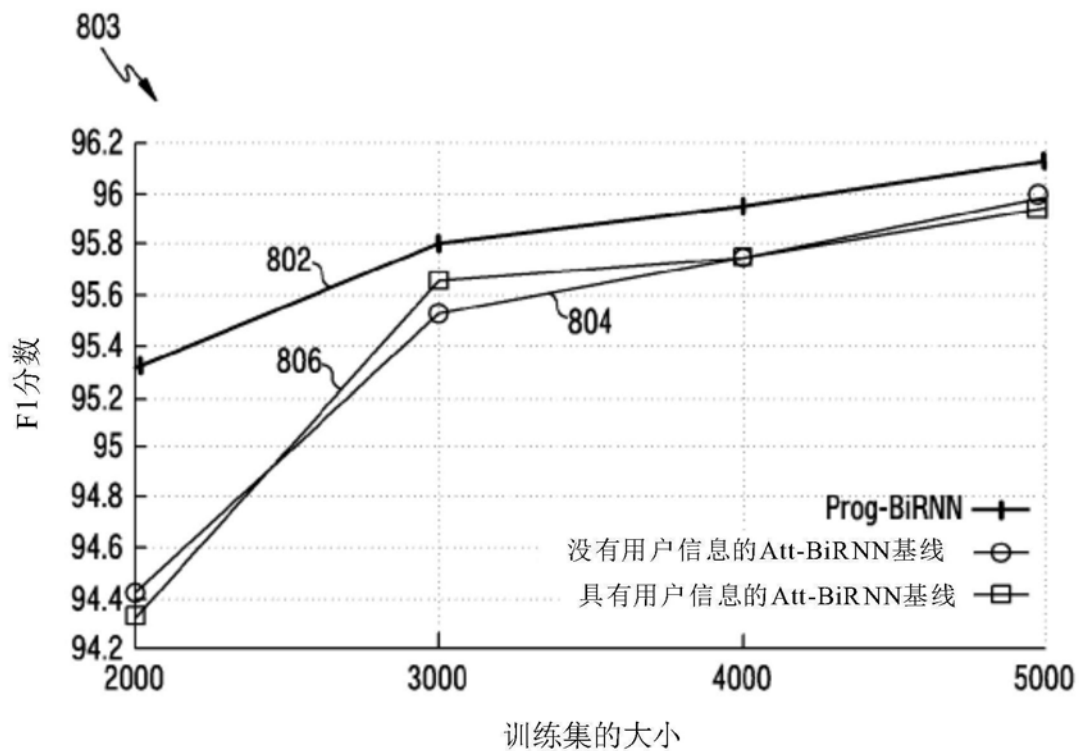


图8B

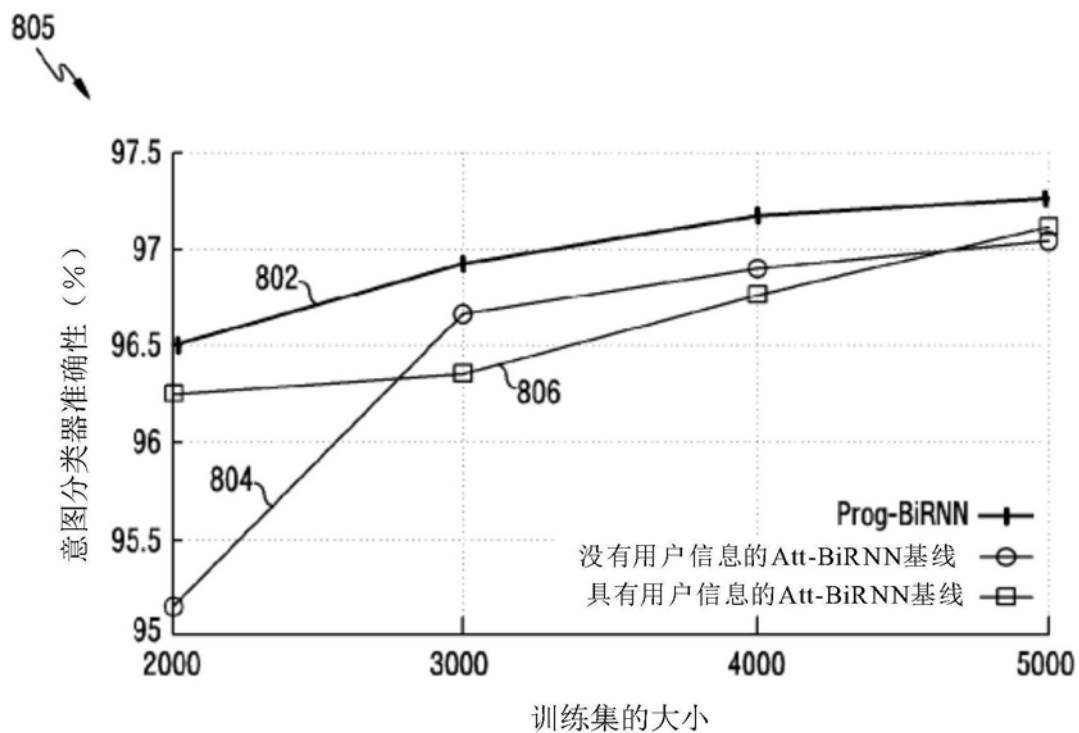


图8C

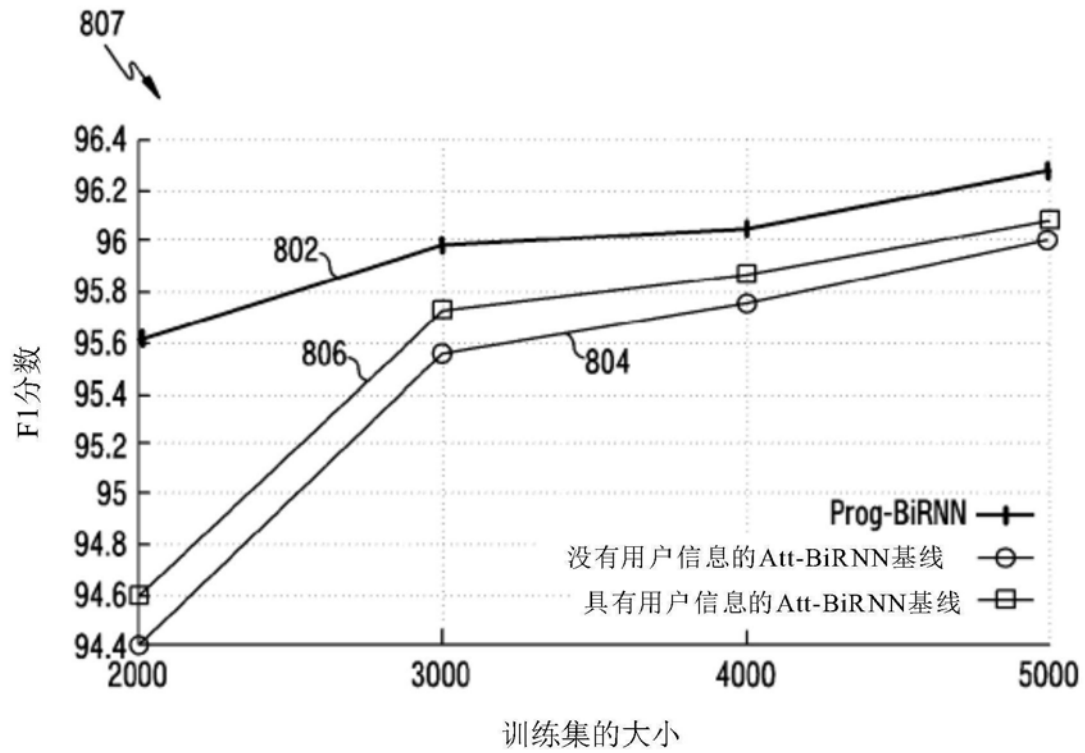


图8D

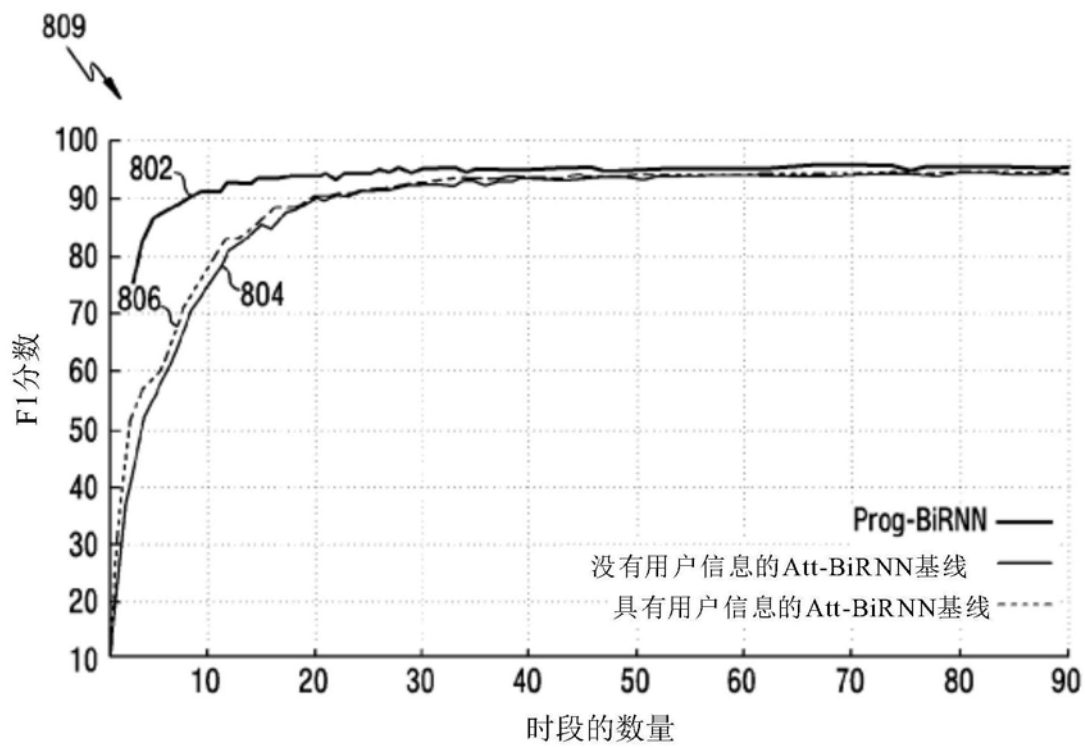


图8E