(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112567394 B
(45)授权公告日 2025.04.15

(21)申请号 201980053458.4
(22)申请日 2019.08.16
(65)同一申请的已公布的文献号
申请公布号 CN 112567394 A
(43)申请公布日 2021.03.26
(30)优先权数据
62/765,005 2018.08.16 US
16/542,017 2019.08.15 US
(85)PCT国际申请进入国家阶段日
2021.02.09
(86)PCT国际申请的申请数据
PCT/US2019/046841 2019.08.16
(87)PCT国际申请的公布数据
WO2020/037217 EN 2020.02.20
(73)专利权人 甲骨文国际公司
地址 美国加利福尼亚
(72)发明人 G·辛拉朱 P·V·安马纳布罗卢
(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
专利代理师 邹丹
(51)Int.Cl.
G06N 5/022(2023.01)
G06N 5/04(2023.01)
G06N 20/00(2019.01)
G06N 3/006(2023.01)
G06N 3/08(2023.01)
G06F 16/955(2019.01)
G06F 18/24(2023.01)
G06F 40/295(2020.01)
G06F 16/36(2019.01)
(56)对比文件
CN 106462608 A,2017.02.22
CN 107704480 A,2018.02.16
审查员 夏冰

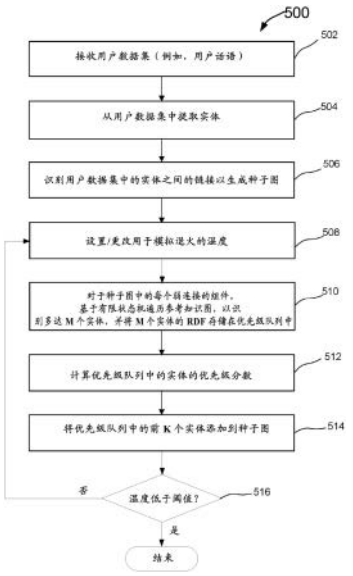
权利要求书4页 说明书38页 附图13页

(54)发明名称

用于在有限的知识领域中构建知识图的技术

(57)摘要

本文公开的技术一般而言涉及构造定制知识图。在一个实施例中,基于某些规则从用户数据集中提取实体以及实体之间的关系以生成种子图。然后使用有限状态机遍历大型知识图,以识别要添加到种子图的候选实体和/或关系。优先级函数用于从候选实体和/或关系中选择实体和/或关系。然后将所选择的实体和/或关系添加到种子图,以生成定制知识图。



CN 112567394 B

1.一种用于为应用生成定制知识图的计算机实现的方法,所述计算机实现的方法包括:

接收应用的用户数据集;

从用户数据集中提取实体;

基于用户数据集识别实体之间的链接;

创建表示实体以及实体之间的链接的种子图,所述种子图包含一组节点和一组边,所述一组节点中的每个节点表示从用户数据集中提取的实体,所述一组边中的每个边连接两个节点并且表示由所述两个节点表示的实体之间的链接;

在种子图中识别一个或多个弱连接的组件,其中每个弱连接的组件是种子图的子图,在所述子图中能够通过沿着无向路径遍历所述子图的边来到达所述子图中的所有节点;

对于种子图中每个弱连接的组件:

将弱连接的组件中的实体映射到参考知识图中的顶点;

从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及

将识别出的实体的资源描述框架RDF保存为缓冲区中的条目;

计算缓冲区中的条目的优先级分数;

从缓冲区中的条目中选择具有最高优先级分数的第一组条目;以及

将由第一组条目识别的实体和链接添加到种子图,以生成定制知识图的版本。

2.如权利要求1所述的计算机实现的方法,其中用户数据集包括多个用户话语。

3.如权利要求1所述的计算机实现的方法,其中提取实体并识别实体之间的链接包括在用户数据集上执行以下中的至少一项:

词性标记;

命名实体辨识;或者

构成解析。

4.如权利要求1所述的计算机实现的方法,其中将弱连接的组件中的实体映射到参考知识图中的顶点包括使用Dexter 2将实体映射到参考知识图中的顶点。

5.如权利要求1所述的计算机实现的方法,其中:

有限状态机包括消歧状态、实体状态、禁止状态和结束状态;以及

遍历参考知识图包括:

当参考知识图中的下一个顶点为消歧顶点时,进入消歧状态;

当参考知识图中的下一个顶点为禁止顶点时:

如果有限状态机的当前状态不是禁止状态,那么进入禁止状态;

当参考知识图中的下一个顶点为实体顶点时:

进入实体状态;以及

将参考知识图中的下一个顶点的RDF保存为缓冲区中的条目;以及

当缓冲区中的条目的数量大于第二阈值数量时,进入结束状态。

6.如权利要求5所述的计算机实现的方法,其中禁止顶点包括日期、时间值、到另一个参考知识图的链接、网页或重复或重定向的实体。

7.如权利要求1所述的计算机实现的方法,其中使用优先级函数来确定缓冲区中的条

目的优先级分数,所述优先级函数是以下各项的函数:

包括种子图和与条目相关联的实体的知识图中的弱连接的组件的数量;

包括种子图和与条目相关联的实体的知识图的图密度;以及

控制参数。

8.如权利要求7所述的计算机实现的方法,还包括:在将在第一组条目中识别的实体和链接添加到种子图之后,更新控制参数。

9.如权利要求8所述的计算机实现的方法,其中更新控制参数包括将控制参数乘以小于1的值。

10.如权利要求8所述的计算机实现的方法,还包括:

对于定制知识图的版本中的每个弱连接的组件:

将弱连接的组件中的实体映射到参考知识图中的顶点;

从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及

将识别出的实体的RDF保存为缓冲区中的条目;

使用优先级函数和更新的控制参数来计算缓冲区中的条目的优先级分数;

从缓冲区中的条目中选择具有最高优先级分数的第二组条目;以及

将由第二组条目识别的实体和链接添加到种子图,以生成更新的定制知识图。

11.如权利要求1所述的计算机实现的方法,还包括,在计算缓冲区中的条目的优先级分数之前,对缓冲区中的条目执行蓄水池抽样以减少缓冲区中的条目的数量。

12.如权利要求1所述的计算机实现的方法,还包括:

接收输入话语;

从输入话语中提取实体;

基于定制的知识图为提取出的实体生成知识图嵌入;以及

基于提取出的实体的知识图嵌入对输入话语进行分类。

13.如权利要求12所述的计算机实现的方法,其中生成知识图嵌入包括使用TransE或TransE的扩展来生成知识图嵌入。

14.如权利要求12所述的计算机实现的方法,其中生成知识图嵌入包括使用TransE和加权评分函数来生成知识图嵌入。

15.如权利要求12所述的计算机实现的方法,其中生成知识图嵌入包括:

使用参考知识图对从输入话语中提取出的实体的知识图嵌入进行预训练;以及

使用定制知识图和预训练的知识图嵌入对从输入话语中提取出的实体的知识图嵌入进行重新训练。

16.如权利要求12所述的计算机实现的方法,其中基于提取出的实体的知识图嵌入对输入话语进行分类包括:

确定提取出的实体的知识图嵌入的平均知识图嵌入;

确定输入话语中的词的GloVe向量的平均GloVe向量;以及

基于平均知识图嵌入和平均GloVe向量对输入话语进行分类。

17.如权利要求16所述的计算机实现的方法,其中基于平均知识图嵌入和平均GloVe向量对输入话语进行分类包括:

组合平均知识图嵌入和平均GloVe向量以生成级联向量;

对级联向量执行主组件分析;以及

基于主组件分析对输入话语进行分类。

18.如权利要求16所述的计算机实现的方法,其中确定输入话语中的词的GloVe向量的平均GloVe向量包括:

清洗输入话语;

为清洗后的输入话语中的每个词生成GloVe向量;以及

对清洗后的输入话语中的每个词的GloVe向量求平均。

19.一种非暂态计算机可读介质,存储能够由一个或多个处理器执行的多条指令,其中所述多条指令当由所述一个或多个处理器执行时,使所述一个或多个处理器执行包括以下各项的处理:

接收应用的用户数据集;

从用户数据集中提取实体;

基于用户数据集识别实体之间的链接;

创建表示实体以及实体之间的链接的种子图,所述种子图包含一组节点和一组边,所述一组节点中的每个节点表示从用户数据集中提取的实体,所述一组边中的每个边连接两个节点并且表示由所述两个节点表示的实体之间的链接;

在种子图中识别一个或多个弱连接的组件,其中每个弱连接的组件是种子图的子图,在所述子图中能够通过沿着无向路径遍历所述子图的边来到达所述子图中的所有节点;

对于种子图中的每个弱连接的组件:

将弱连接的组件中的实体映射到参考知识图中的顶点;

从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及

将识别出的实体的资源描述框架RDF保存为缓冲区中的条目;

计算缓冲区中的条目的优先级分数;

从缓冲区中的条目中选择具有最高优先级分数的第一组条目;以及

将在第一组条目中识别的实体和链接添加到种子图,以生成定制知识图的版本。

20.一种用于为应用生成定制的知识图的系统,所述系统包括:

一个或多个处理器;以及

耦合到所述一个或多个处理器的存储器,所述存储器存储指令,所述指令当由所述一个或多个处理器执行时,使所述系统执行包括以下各项的操作:

接收应用的用户数据集;

从用户数据集中提取实体;

基于用户数据集识别实体之间的链接;

创建表示实体以及实体之间的链接的种子图,所述种子图包含一组节点和一组边,所述一组节点中的每个节点表示从用户数据集中提取的实体,所述一组边中的每个边连接两个节点并且表示由所述两个节点表示的实体之间的链接;

在种子图中识别一个或多个弱连接的组件,其中每个弱连接的组件是种子图的子图,在所述子图中能够通过沿着无向路径遍历所述子图的边来到达所述子图中的所有节点;

对于种子图中每个弱连接的组件：

将弱连接的组件中的实体映射到参考知识图中的顶点；

从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及

将识别出的实体的资源描述框架RDF保存为缓冲区中的条目；

计算缓冲区中的条目的优先级分数；

从缓冲区中的条目中选择具有最高优先级分数的第一组条目;以及

将在第一组条目中识别的实体和链接添加到种子图,以生成定制知识图的版本。

用于在有限的知识领域中构建知识图的技术

[0001] 相关申请的交叉引用

[0002] 本申请要求于2018年8月16日提交的题为“Techniques For Building A Knowledge Graph In Limited Knowledge Domains”的美国临时专利申请号62/765,005和于2019年8月15日提交的题为“TECHNIQUES FOR BUILDING A KNOWLEDGE GRAPH IN LIMITED KNOWLEDGE DOMAINS”的美国专利申请No.16/542,017的优先权权益,这些申请的包括任何附录的全部公开内容为了所有目的通过引用并入本文。

背景技术

[0003] 知识图通过明确描述实体之间的关系以结构化的方式组织信息。知识图通常使用有向图来表示知识库,该有向图在节点或顶点(表示实体)之间具有带标记的边(表示关系)。知识图可以用于例如搜索、问题回答、对话接口(例如,聊天机器人)、推荐系统等。有一些大规模而全面的知识图可用。但是,用于特定应用或环境的自然语言理解(NLU)服务(诸如,为不同用户构建对话接口(例如,聊天机器人))通常在特定领域内表现更好。为特定知识领域创建准确的知识图是一项具有挑战性的任务,部分原因是自然语言的微妙和歧义性以及特定领域中由于用户数据集中样本数量通常很小而导致知识有限。

发明内容

[0004] 本公开一般而言涉及构建定制的知识图,并且更特别地涉及用于基于小的用户数据集和大规模参考(例如,外部)知识图为特定知识领域(例如,为特定客户端或特定机器人)构造定制的知识图的技术。定制的知识图可以用于例如基于知识图嵌入技术来改善聊天机器人中的意图分类。本文描述了各种发明实施例,包括方法、系统、存储由一个或多个处理器可执行的程序、代码或指令的非暂态计算机可读存储介质等。

[0005] 在某些实施例中,公开了用于为应用生成定制的知识图的计算机实现的方法。该方法可以包括:接收应用的用户数据集;从用户数据集中提取实体;基于用户数据集识别实体之间的链接;以及创建表示实体和实体之间的链接的种子图。该方法还可以包括:识别种子图中弱连接的组件,并且对于种子图中每个弱连接的组件,将弱连接的组件中的实体映射到参考知识图中的顶点;从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及将识别出的实体的资源描述框架(RDF)保存为缓冲区中的条目。该方法还可以包括:计算缓冲区中的条目的优先级分数;从缓冲区中的条目中选择具有最高优先级分数的第一组条目;以及将由第一组条目识别的实体和链接添加到种子图以生成定制知识图的版本。

[0006] 在一些实施例中,用户数据集可以包括多个用户话语。提取实体并识别实体之间的链接可以包括例如对用户数据集执行词性(part-of-speech)标记、命名实体辨识和/或构成(constituency)解析。将弱连接的组件中的实体映射到参考知识图中的顶点可以包括使用例如Dexter 2将实体映射到参考知识图中的顶点。

[0007] 在一些实施例中,有限状态机可以包括消歧状态、实体状态、禁止状态和结束状

态。遍历参考知识图可以包括:当参考知识图中的下一个顶点为消歧顶点时,进入消歧状态;当参考知识图中的下一个顶点为禁止顶点时,并且如果有限状态机的当前状态不是禁止状态,那么进入禁止状态;当参考知识图中的下一顶点为实体顶点时,进入实体状态,并将参考知识图中的下一顶点的RDF保存为缓冲区中的条目;以及当缓冲区中的条目的数量大于第二阈值数量时,进入结束状态。禁止顶点可以包括例如日期、时间值、到另一个参考知识图的链接、网页或重复或重定向的实体。

[0008] 在一些实施例中,缓冲区中的条目的优先级分数可以使用优先级函数来确定,该优先级函数是包括种子图和与条目相关联的实体的知识图中的弱连接的组件的数量、包括种子图和与条目相关联的实体的知识图的图密度和/或控制参数的函数。在一些实施例中,该方法还可以包括:在为缓冲区中的条目计算优先级分数之前,对缓冲区中的条目执行蓄水池(reservoir)抽样以减少缓冲区中的条目的数量。

[0009] 在一些实施例中,该方法还可以包括在将在第一组条目中识别的实体和链接添加到种子图之后,更新控制参数。更新控制参数可以包括将控制参数乘以小于1的值。在一些实施例中,该方法还可以包括,对于定制知识图的版本中的每个弱连接的组件,将弱连接的组件中的实体映射到参考知识图中的顶点;从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及将识别出的实体的RDF保存为缓冲区中的条目。该方法还可以包括:使用优先级函数和更新的控制参数来计算缓冲区中的条目的优先级分数;从缓冲区中的条目中选择具有最高优先级分数的第二组条目;以及将由第二组条目识别的实体和链接添加到种子图以生成更新的定制知识图。

[0010] 在一些实施例中,该方法还可以包括:接收输入话语、从输入话语中提取实体、基于定制的知识图为提取出的实体生成知识图嵌入,以及基于提取出的实体的知识图嵌入对输入话语进行分类。在一些实施例中,生成知识图嵌入可以包括使用TransE或TransE的扩展来生成知识图嵌入。在一些实施例中,生成知识图嵌入可以包括使用TransE和加权评分函数来生成知识图嵌入。在一些实施例中,生成知识图嵌入可以包括:使用参考知识图对从输入话语提取出的实体的知识图嵌入进行预训练;以及使用定制知识图和预训练的知识图嵌入对从输入话语提取出的实体的知识图嵌入进行重新训练。

[0011] 在一些实施例中,基于提取出的实体的知识图嵌入对输入话语进行分类可以包括:确定提取出的实体的知识图嵌入的平均知识图嵌入;确定输入话语中的词的GloVe向量的平均GloVe向量;以及基于平均知识图嵌入和平均GloVe向量对输入话语进行分类。在一些实施例中,基于平均知识图嵌入和平均GloVe向量对输入话语进行分类可以包括:组合平均知识图嵌入和平均GloVe向量以生成级联向量;对级联向量执行主组件分析;以及基于主组件分析对输入话语进行分类。在一些实施例中,确定输入话语中的词的GloVe向量的平均GloVe向量可以包括:清洗输入话语;为清洗后的输入话语中的每个词生成GloVe向量;以及对清洗后的输入话语中的每个词的GloVe向量求平均。

[0012] 根据某些实施例,一种非暂态计算机可读介质可以存储由一个或多个处理器可执行的多条指令,其中该多条指令当由该一个或多个处理器执行时,可以使该一个或多个处理器执行处理,包括:接收应用的用户数据集;从用户数据集中提取实体;基于用户数据集识别实体之间的链接;以及创建表示实体以及实体之间的链接的种子图。该处理还可以包

括:在种子图中识别弱连接的组件,并且对于种子图中每个弱连接的组件,将弱连接的组件中的实体映射到参考知识图中的顶点;从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及将识别出的实体的资源描述框架(RDF)保存为缓冲区中的条目。该处理还可以包括:计算缓冲区中的条目的优先级分数;从缓冲区中的条目中选择具有最高优先级分数的第一组条目;以及将由第一组条目识别的实体和链接添加到种子图,以生成定制知识图的版本。

[0013] 根据某些实施例,一种系统可以包括一个或多个处理器以及耦合到该一个或多个处理器的存储器。存储器可以存储指令,该指令当被一个或多个处理器执行时,可以使系统执行操作,包括:接收应用的用户数据集;从用户数据集中提取实体;基于用户数据集识别实体之间的链接;以及创建表示实体以及实体之间的链接的种子图。该操作还可以包括:在种子图中识别弱连接的组件,并且对于种子图中每个弱连接的组件,将弱连接的组件中的实体映射到参考知识图中的顶点;从参考知识图中被映射的顶点开始并基于有限状态机遍历参考知识图,以识别参考知识图中多达第一阈值数量的实体;以及将识别出的实体的资源描述框架(RDF)保存为缓冲区中的条目。该操作还可以包括:计算缓冲区中的条目的优先级分数;从缓冲区中的条目中选择具有最高优先级分数的第一组条目;以及将由第一组条目识别的实体和链接添加到种子图,以生成定制知识图的版本。

[0014] 通过参考以下说明书、权利要求书和附图,前述以及其它特征和实施例将变得更加清楚。

附图说明

[0015] 下面参考以下附图详细描述说明性示例。

[0016] 图1描绘了根据某些实施例的分布式系统,该分布式系统实现用于使用消息传递应用与最终用户进行通信的机器人系统。

[0017] 图2描绘了根据某些实施例的包括机器人系统和机器人分析系统的集成系统,该机器人分析系统用于监视、分析、可视化和改善机器人系统的性能。

[0018] 图3图示了知识图的示例。

[0019] 图4A-4D图示了构建知识图的处理的示例。

[0020] 图5是图示根据某些实施例的用于为特定领域(例如,为聊天机器人)构建定制知识图的方法的示例的简化流程图。

[0021] 图6图示了根据某些实施例的用于控制参考知识图的遍历的有限状态机(FSM)的示例。

[0022] 图7图示了根据某些实施例的基于用户数据集(例如,用户话语)生成的种子图的示例。

[0023] 图8图示了根据某些实施例的基于图7中所示的种子图生成的定制知识图的示例。

[0024] 图9是图8中绘出的定制知识图的一部分的放大图。

[0025] 图10描绘了用于实现某些实施例的分布式系统的简化图。

[0026] 图11是根据某些实施例的其中各种服务可以被提供为云服务的基于云的系统环境的简化框图。

[0027] 图12图示了可以用于实现某些实施例的计算机系统的示例。

具体实施方式

[0028] 本公开一般而言涉及构建定制的知识图,并且更特别地涉及用于使用领域中的有限知识(例如,小的用户数据集)和大规模参考(例如,外部)知识图为特定知识领域构造定制的知识图的技术。定制的知识图用于例如基于知识图嵌入技术来改善聊天机器人中的意图分类。本文描述了各种发明实施例,包括方法、系统、存储由一个或多个处理器可执行的程序、代码或指令的非暂态计算机可读存储介质等。

[0029] 用于特定应用或环境的自然语言理解(NLU)服务,诸如针对不同用户的对话接口(例如,聊天机器人),可以使用特定领域内的定制知识图而不是大的综合知识图来更好地执行。例如,定制知识图中的噪声可能更小,因此使用小的定制知识图而不是大的通用知识图可能更高效并且更准确。但是,特定领域中的可用知识可能是有限的,因此基于有限知识生成的定制知识图可能是稀疏的。

[0030] 根据一些实施例,可以基于某些规则从小的用户数据集(包括,例如,用户话语、意图、实体和QnA)提取用于自定义应用(例如,聊天机器人)的实体。可以基于提取出的实体以及实体之间的连接或关系来生成种子图。然后可以使用有限状态机遍历大规模参考知识图(诸如WikiData)来识别要添加到种子图的候选实体和/或关系,以将种子图扩展为用于自定义应用的定制知识图。遍历还可以帮助从参考知识图中识别用户数据集中的实体与相关事实之间的可能附加关系。评分功能可以用于从识别出的实体和/或关系中选择实体和/或关系,然后将所选择的实体和/或关系添加到种子图以生成定制的知识图。使用某些知识图嵌入技术,可以将生成的定制知识图用于客户应用,诸如自定义聊天机器人。消融(ablation)研究表明,定制知识图捕获的附加语义信息可以提高自定义聊天机器人的性能。本文公开的技术还可以在其它应用中使用,诸如问题回答或关系提取。

[0031] 如本文所使用的,“聊天机器人”、“机器人”或“技能”是指被设计为模拟尤其是通过互联网与人类用户对话的计算机程序。可以设计个体技能来与用户交互并完成特定类型的任务,诸如订购食物、进行预订和更改联系信息。每种技能可以通过组合文本消息和简单的UI元素(如选择列表)来帮助用户完成任务。

[0032] 如本文所使用的,术语“意图”可以指用户期望技能为其执行的动作或任务的类别。术语“实体”可以指从用户输入中识别使技能能够完成任务的关键信息的变量。术语“组件”可以指技能可以用来响应用户的各种功能。术语“组件”可以指通用功能,诸如输出文本或从后端返回信息和执行自定义逻辑。术语“对话流”可以指技能-用户交互的定义并且可以描述技能如何根据用户输入进行响应和表现。术语“通道”可以指特定于平台的配置,以允许技能访问消息传递平台或客户端消息传递应用。单个技能可能具有为其配置的若干个通道,使得它可以同时在不同的服务或平台上运行。

[0033] 如本文所使用的,话语或消息可以指在对话期间交换的一个或多个句子,其中对话可以指可以包括一个或多个话语或消息的通信会话。对话可以包括一个或多个阶段或状态。对话流可以是包括相同阶段或状态以及从阶段(或状态)到阶段(或状态)的相同转换的多个对话的抽象。每个对话可以是对应对话流的具体实例。对话或对话流的状态或阶段可以与由机器人系统维护的用于与其它机器人系统或人进行对话的状态机的状态相关联。在一些情况下,状态可以对应于最终用户的意图或目标。如本文所使用的,最终用户可以指机器人系统的最终用户,诸如通过消息传递应用或平台与机器人系统进行对话的人或另一个

实体。例如,最终用户可以是拥有机器人系统的企业的客户或客户端。如本文所使用的,分析系统的用户可以指机器人系统的所有者、操作员、管理员或开发人员。

[0034] 在以下描述中,出于解释的目的,阐述了具体细节以便提供对本公开的示例的透彻理解。但是,将清楚的是,可以在没有这些具体细节的情况下实践各种示例。随后的描述仅提供示例,并且无意于限制本公开的范围、适用性或配置。相反,随后的示例描述将向本领域技术人员提供用于实现示例的可行描述。应该理解的是,在不脱离所附权利要求书所阐述的本公开的精神和范围的情况下,可以对元件的功能和布置进行各种改变。各图和描述不意图是限制性的。电路、系统、网络、处理和其它组件可能以框图形式示出为组件,以便不会在不必要的细节上使示例模糊。在其它情况下,可能示出众所周知的电路、处理、算法、结构和技术而没有不必要的细节,以避免使示例模糊。本文公开的教导还可以应用于各种类型的应用,诸如移动应用、非移动应用、桌面应用、web应用、企业应用等。此外,本公开的教导不限于特定的操作环境(例如,操作系统、设备、平台等),而是可以应用于多个不同的操作环境。

[0035] 另外,注意的是,各个示例可以被描述为处理,该处理被描绘为流程图、流图、数据流图、结构图或框图。虽然流程图可以将操作描述为顺序处理,但是许多操作可以并行或同时执行。另外,可以重新布置操作的顺序。当处理的操作完成时处理会终止,但是可能有图中没有包括的附加步骤。处理可以对应于方法、函数、过程、子例程、子程序等。当处理对应于函数时,其终止可以对应于该函数返回到调用函数或主函数。

[0036] 词语“示例”或“示例性”在本文中用来表示“用作示例、实例或说明”。本文中描述为“示例性”或“示例”的任何实施例或设计不一定被解释为比其它实施例或设计更优选或有利。

[0037] 术语“机器可读存储介质”或“计算机可读存储介质”包括但不限于便携式或非便携式存储设备、光学存储设备以及能够存储、包含、或携带(一条或多条)指令和/或数据的各种其它介质。机器可读存储介质或计算机可读存储介质可以包括非暂态介质,其中可以存储数据并且不包括无线或通过有线连接传播的载波和/或暂态电子信号。非暂态介质的示例可以包括但不限于磁盘或磁带、诸如光盘(CD)或数字通用盘(DVD)的光学存储介质、闪存、存储器或存储器设备。计算机程序产品可以包括代码和/或机器可执行指令,其可以表示过程、函数、子程序、程序、例程、子例程、模块、软件包、类,或指令、数据结构或程序语句的任意组合。可以通过传递和/或接收信息、数据、自变量、参数或存储器内容,将代码段耦合到另一个代码段或硬件电路。信息、自变量、参数、数据等可以经由任何合适的手段传递、转发或传输,包括存储器共享、消息传递、令牌传递、网络传输等。

[0038] 此外,可以通过硬件、软件、固件、中间件、微代码、硬件描述语言或其任意组合来实现示例。当以软件、固件、中间件或微代码实现时,用于执行必要任务的程序代码或代码段(例如,计算机程序产品)可以存储在机器可读介质中。(一个或多个)处理器可以执行必要的任务。

[0039] 一些图中描绘的系统可以以各种配置来提供。在一些示例中,系统可以被配置为分布式系统,其中系统的一个或多个组件被分布在云计算系统中的一个或多个网络上。

[0040] 在将组件描述为“被配置为”执行某些操作的情况下,可以例如通过设计电子电路或其它硬件来执行操作、通过对电子电路(例如,微处理器或其它合适的电子电路)进行编

程或控制来执行操作或其任意组合来完成这样的配置。

[0041] 为了获得即时反应,世界各地的许多用户都在即时消息传递或聊天平台上。组织经常使用这些即时消息传递或聊天平台与客户(或最终用户)进行实时对话。但是,组织雇用服务人员与客户或最终用户进行实时通信可能会非常昂贵。因此,已经开始开发聊天机器人(也称为机器人(bot)、技能、聊天机器人(chatterbot)或对话(talkbot)机器人)来模拟与最终用户的对话,尤其是通过互联网。最终用户可以通过最终用户已经安装和使用的消息传递应用与机器人通信。通常由人工智能(AI)驱动的智能机器人可以在实时对话中进行更智能、更上下文感知的通信,因此可以使机器人与最终用户之间的对话更自然,从而改善对话体验。不是最终用户学习机器人知道如何响应的一组固定的关键字或命令,而是智能机器人可以能够基于自然语言中的用户话语来理解最终用户的意图并做出相应的响应。

[0042] 机器人系统可以包括:一个或多个用户意图分类引擎,用于基于最终用户话语识别最终用户意图;以及一个或多个对话引擎,用于根据所确定的最终用户意图对用户话语进行智能和上下文感知响应。但是,构建机器人系统(诸如,可以基于用户话语确定最终用户意图的用户意图分类引擎)是一项具有挑战性的任务,部分原因是自然语言的微妙和歧义性以及输入空间(例如,可能的用户话语)的维度和输出空间的大小(意图的数量)。

[0043] 机器人通常可以通过使用自然语言消息的消息传递应用来响应自然语言消息(例如,问题或评论)。企业可以使用一个或多个机器人系统通过消息传递应用与最终用户通信。可以被称为通道的消息传递应用可以是最终用户已经安装并熟悉的最终用户优选的消息传递应用。因此,最终用户无需下载和安装新应用即可与机器人系统聊天。消息传递应用可以包括例如上层(OTT)消息传递通道(诸如,Facebook Messenger、Facebook WhatsApp、微信、Line、Kik、Telegram、Talk、Skype、Slack或SMS)、虚拟私人助理(诸如,Amazon Dot、Echo或Show、Google Home、Apple HomePod等)、扩展为具有聊天能力的本地或混合/响应移动应用或web应用的移动和web应用扩展、或基于语音的输入(诸如,具有使用Siri、Cortana、Google Voice或其它言语输入进行交互的接口的设备或应用)。

[0044] 在一些示例中,机器人系统可以与统一资源标识符(URI)相关联。URI可以使用字符串来识别机器人系统。URI可以用作一个或多个消息传递应用系统的webhook(网络钩子)。URI可以包括例如统一资源定位符(URL)或统一资源名称(URN)。可以将机器人系统设计为从消息传递应用系统接收消息(例如,超文本传输协议(HTTP)post调用消息)。可以从消息传递应用系统将HTTP post调用消息定向到URI。在一些实施例中,该消息可以不同于HTTP post调用消息。例如,机器人系统可以从短消息服务(SMS)接收消息。虽然本文的讨论可能指机器人系统接收到的作为消息的通信,但是本领域普通技术人员将认识到消息可以是HTTP post调用消息、SMS消息或两个系统之间的任何其它类型的通信。

[0045] 最终用户可以通过对话交互(有时称为对话用户接口(UI))与机器人系统进行交互,就像人与人之间的交互一样。在一些情况下,交互可以包括最终用户向机器人说“你好(Hello)”,并且机器人以“嗨(Hi)”响应并询问最终用户如何提供帮助。在一些情况下,交互也可以是与例如银行机器人的交易交互,诸如将钱从一个账户转移到另一个账户;与例如HR机器人的信息交互,诸如检查假期剩余天数;或与例如零售机器人的交互,诸如讨论退货或寻求技术支持。

[0046] 在一些实施例中,机器人系统可以智能地应对最终用户交互,而无需与机器人系

统的管理员或开发人员进行交互。例如,最终用户可以向机器人系统发送一条或多条消息,以实现期望的目标。消息可以包括某些内容,诸如文本、表情符号、音频、图像、视频或其它传达消息的方法。在一些实施例中,机器人系统可以将内容转换成标准化形式(例如,针对具有适当参数的企业服务的代表性状态转移(REST)调用)并生成自然语言响应。机器人系统还可以提示最终用户提供附加输入参数或请求其它附加信息。在一些实施例中,机器人系统还可以发起与最终用户的通信,而不是被动地响应最终用户话语。

[0047] 与机器人的对话可以遵循包括多个状态的特定对话流。该流可以基于输入来定义接下来将发生什么。在一些实施例中,包括用户定义的状态(例如,最终用户意图)和在状态中或从一个状态到另一个状态要采取的动作的状态机可以用于实现机器人系统。对话可以基于最终用户输入采用不同的路径,这可能会影响机器人针对该流做出的决定。例如,在每种状态下,机器人都可以基于最终用户输入来确定最终用户的意图,以便确定要采取的下一个适当的动作。

[0048] 意图可以包括最终用户想要实现的目标。意图将最终用户输入映射到后端系统应为最终用户执行的动作。因此,基于最终用户以自然语言说出的短语,机器人会将最终用户话语映射到具体用例或工作,诸如订购比萨、获取账户余额、转移资金、进行购买、退货等。人类对话通常本质上是非线性的。最终用户在对话期间可能经常会进入不同的状态。例如,如果最终用户想要将资金从账户A转移到接收者,那么最终用户可以通过例如要求机器人向接收者支付晚餐费用开始与机器人系统进行对话。机器人可以用“使用哪个账户?(from which account?)”进行响应。最终用户可以选取支票账户,但随后可能意识到他或她不确定该账户中的余额。因此,最终用户可能切换上下文来询问余额和最近的交易,等等。换句话说,最终用户可能会触发流和状态的更改,例如,从转移资金到检查余额,然后再到最近的交易。在某个时间点,最终用户可能决定返回到原始意图-向接收者付款。因此,机器人系统的一项任务是从自然语言话语中动态确定最终用户的意图。

[0049] 机器人可以使用自然语言处理(NLP)引擎和/或机器学习模型(例如,意图分类器)将最终用户的话语映射到具体意图。例如,基于机器学习的NLP引擎可以学习理解和分类来自最终用户的自然语言对话,并从对话中提取必要的信息,以便能够采取精确的动作,诸如执行交易或从后端记录系统中查找数据。

[0050] 图1描绘了根据某些实施例的分布式系统100,其可以用于实现用于使用消息传递应用与最终用户通信的机器人系统。系统100可以包括机器人系统120、一个或多个消息传递应用系统115以及一个或多个最终用户设备,诸如一个或多个移动设备110。在一些示例中,消息传递应用可以安装在电子设备上(例如,台式计算机、膝上型电脑、移动设备110等)。虽然本文的讨论将参考移动设备和消息传递应用,但是本领域普通技术人员将认识到可以使用任何电子设备,并且可以使用任何消息传递平台或消息传递应用,诸如FACEBOOK® Messenger、WHATSAPP®即时消息传递软件、WECHAT®移动文本和语音消息传递通信服务、KIK® Messenger、TELEGRAM® Messenger、SKYPEMOBILE® Messenger、短消息服务(SMS)或任何其它为最终用户提供用于通信的平台的消息传递应用。在其它示例中,可以通过安装在移动设备110上的浏览器(例如,GOOGLECHROME®浏览器、MOZILLA® FIREFOX®浏览器和INTERNET

EXPLORER浏览器)来运行消息传递应用。在一些实施例中,两个或更多个消息传递应用可以安装在最终用户设备上,以通过两个或更多个消息传递平台(诸如,两个或更多个消息传递应用系统115)进行通信。

[0051] 可以通过诸如消息传递应用系统115之类的消息传递平台来促进消息传递应用。移动设备110可以通过第一网络(例如,互联网)连接到消息传递应用系统115。消息传递应用系统115可以是由诸如Facebook、腾讯、Google、Microsoft等第三方提供的消息传递平台。消息传递应用系统115可以管理通过消息传递应用在多个移动设备或其它最终用户设备上发送和接收的内容。

[0052] (例如,在一个或多个服务器上实现的)机器人系统120也可以通信地连接到消息传递应用系统115,以发送和接收消息。消息传递应用系统115和机器人系统120之间的通信可以通过第二网络(例如,互联网)。第一网络和第二网络可以是相同的网络,或者它们可以是相似或完全不同的网络。消息传递应用系统115可以使用互联网将内容(例如,消息或消息中的信息)从移动设备110路由到机器人系统120。在一些实施例中,内容的目的地(例如,机器人系统120的标识)可以被包括在内容中作为名义被访者(addressee)。在一些实施例中,机器人系统120还可以被配置为与两个或更多个消息传递应用系统115通信。

[0053] 如上所述,在最终用户之间或最终用户与机器人系统之间交换的内容可以包括例如文本、表情符号、音频、媒体(例如,图片、视频、链接)或传达消息的任何其它方法。机器人系统120从例如**FACEBOOK®** Messenger接收到的消息的示例可以包括:

[0054]

```
{
    "channel": {
        "tenant_id": "DemoTenant",
        “client_data": {
            "channel_user_id": "1103645260000000",
            "type": "facebook",
            "chat_id": "1103645260000000"
        },
        "type": "facebook",
        "bot_id": "DemoBot",
        "client_id": "facebook"
    },
    "id": "411fc848-350b-47de-89c3-a0ecee314401",
    "text": "May I have a cheese pizza?",
    "type": "text"
}
```

[0055] 机器人系统120可以使用连接器130从消息传递应用系统115接收内容,该连接器130充当消息传递应用系统115和机器人系统120之间的接口。在一些实施例中,连接器130可以标准化来自消息传递应用系统115的内容,使得机器人系统120可以跨不同消息传递应用系统分析内容。内容标准化处理可以包括将内容从每种类型的消息传递应用格式化为用于处理的公共格式。在一些实施例中,机器人系统120可以包括用于每种消息传递应用(诸如,FACEBOOK®Messenger、WHATSAPP®即时消息传递软件、WECHAT®移动文本和语音消息传递通信服务、KIK®Messenger、TELEGRAM®Messenger和SKYPEMOBILE®Messenger、短消息服务(SMS))的一个或多个连接器。在一些实施方式中,连接器130可以将内容路由到消息输入(message-in)队列140。消息输入队列140可以包括按接收到的顺序存储内容的缓冲区(例如,先进先出(FIFO)缓冲区)。在一些实施例中,每个连接器130可以与一个或多个消息输入队列相关联。

[0056] 当消息处理器150变得可用时,消息输入队列140可以将内容发送到消息处理器150。在一些实施例中,消息处理器150可以从消息输入队列140中提取内容。消息处理器150可以解析消息并确定所解析的消息的意图,如下面详细描述的。在一些实施例中,消息处理器150可以包括自然语言处理器152和意图确定子系统154。自然语言处理器152可以解析消息并执行某些语义分析,诸如识别主语、谓语(例如,动作)和/或宾语。意图确定子系统154可以基于所解析的消息来确定最终用户意图。如上所述,意图可以包括消息的目的。例如,消息的目的可以是订购比萨、订购计算机、转移资金、询问有关交付的问题等。在一些实施例中,与意图相关联的更具体地定义或阐明要采取的行动的参数(其可以被称为实体)也可以由自然语言处理器152和/或意图确定子系统154从消息中提取。

[0057] 在由消息处理器150基于内容确定最终用户意图之后,可以将所确定的意图(以及与意图相关联的参数)发送给动作引擎160。动作引擎160可以用于基于意图(以及与意图相关联的参数)和状态机的当前状态(或上下文)确定要执行的动作,如上所述。例如,动作引擎160可以将某些出站内容作为响应发送到消息输出(message-out)队列170,和/或可以向诸如企业服务125之类的一些企业服务发送命令或从其检索信息。消息输出队列170可以向连接器130发送出站内容。然后,连接器130可以将出站内容发送到由动作引擎160指示的消息传递应用系统,该系统可以与消息传递应用系统115相同或不同。消息传递应用系统115然后可以将出站内容转发到移动设备110上的消息传递应用。

[0058] 机器人系统120可以与一个或多个企业服务(例如,企业服务125)、用于存储和/或分析由机器人系统120接收到的消息的一个或多个存储系统,或用于向机器人系统120提供内容的内容系统进行通信。企业服务125可以与连接器130、动作引擎160或其任何组合中的一个或多个通信。企业服务125可以以类似于消息传递应用系统115的方式与连接器130通信。企业服务125可以将内容发送到连接器130以与一个或多个最终用户相关联。企业服务125还可以将内容发送到连接器130,以使机器人系统120执行与最终用户关联的动作。动作引擎160可以与企业服务125通信以从企业服务125获得信息和/或指示企业服务125采取由动作引擎160识别的动作。

[0059] 在一些实施例中,机器人系统120可以包括一个或多个计时器。计时器可以使动作引擎160在经过一定时间后使用连接器130和消息传递应用系统115将内容发送给最终用

户。在一些实施例中,计时器可以类似于最终用户或企业服务125向机器人系统120发送内容。例如,因为来自最终用户的消息将被分析,因此计时器可以向机器人系统120发送消息以进行分析。

[0060] 在一个具体实施例中,最终用户可以使用移动设备110通过消息传递应用系统115向机器人系统120发送消息。消息可以包括问候语,诸如“您好(Hello)”或“嗨(Hi)”。机器人系统可以确定与最终用户的新对话已经开始并启动状态机。在一些实施例中,机器人系统可以识别最终用户的一个或多个特点。例如,机器人系统可以使用消息传递应用系统上与最终用户相关联的简档来识别最终用户的名称。使用一个或多个特点,机器人系统可以在消息传递应用上响应最终用户。响应可以包括给最终用户的消息,该消息对从最终用户接收到的消息做出响应。例如,响应中可以包括带有最终用户的名称的问候语,诸如“嗨,Tom,我能为您做什么?(Hi Tom,What can I do for you?)”。取决于与机器人系统相关联的企业,机器人系统可以进展以实现企业的目标。例如,如果机器人系统与比萨递送企业相关联,那么机器人系统可以向最终用户发送消息,询问最终用户是否想要订购比萨。机器人系统与最终用户之间的对话可以从那里继续、来回进行,直到机器人系统已经完成对话或者最终用户停止对机器人系统响应。

[0061] 在一些实施例中,机器人系统可以发起与最终用户的对话。机器人系统发起的对话可以响应于与最终用户的先前对话。例如,最终用户可能在先前对话中订购了比萨。机器人系统然后可以在比萨准备好时发起对话。在一些实施例中,当从与机器人系统相关联的企业接收到指示时(例如,员工向机器人系统发送比萨已经准备好的消息),机器人系统可以确定比萨已经准备好。对话可以包括发送到最终用户的指示比萨已准备好的消息。

[0062] 在一些实施例中,机器人系统可以在与接收到先前消息的消息传递应用不同的消息传递应用上向最终用户发送消息。例如,机器人系统可以确定使用短消息服务(SMS)而不是FACEBOOK®Messenger来发送消息。在这样的实施方式中,机器人系统可以集成多个消息传递应用。

[0063] 在一些实施例中,机器人系统可以基于计时器来确定开始对话。例如,机器人系统可以确定在订购比萨之后为最终用户安排一个星期的计时器。一个星期的计时器到期可以使机器人系统针对订购另一个比萨开始与最终用户的新对话。计时器可以由企业配置并由机器人系统实现。

[0064] 如上所述,在一些实施例中,动作引擎160可以向一些企业服务125发送命令或从其检索信息。例如,当机器人系统120(更具体而言,消息处理器150)确定检查余额的意图时,机器人系统120可以确定若干个账户(例如,支票或储蓄账户)中的哪一个需要检查余额。如果最终用户输入“我的储蓄账户中的余额是多少(What's my balance in my savings account)”,那么机器人系统120可以提取“储蓄(saving)”并将命令发送到银行服务器以检查余额,然后通过消息将接收到的余额信息发送给最终用户。如果最终用户最初仅说出“我的账户中的余额是多少?(what's the balance in my account?)”,那么机器人系统120可以向最终用户发送消息,提示最终用户进一步指定具体账户,或者可以检索最终用户的所有账户的信息并将账户信息发送给最终用户以供最终用户进行选择。

[0065] 在一些实施例中,机器人系统可以维持对话之间的信息。该信息可以在以后使用,使得机器人系统无需每次在最终用户与机器人系统之间开始新对话时都询问一些问题。例

如,机器人系统可以存储与最终用户先前的比萨订购相关的信息。在新对话中,机器人系统可能会向最终用户发送消息,询问最终用户是否想要与上次相同的订购。

[0066] 在一些实施例中,机器人系统120可以将与最终用户相关联的信息存储在高速缓存中。在将出站消息从连接器130发送到消息传递应用系统之后,高速缓存可以写入数据库以保存信息。在其它实施例中,高速缓存可以在不同时间(例如,在特定事件之后、在每个事件之后、在一定时间之后或任何其它确定何时写入数据库的度量)写入数据。

[0067] 当识别出减速时,机器人系统120可以允许缩放每个组件。例如,如果机器人系统120识别到达连接器130的消息数量超过阈值,那么可以向连接器130添加一个或多个附加连接器。此外,取决于减速发生的地方,可以增加消息输入队列、消息处理器、动作引擎实例和消息输出队列的数量。在这样的实施方式中,可以添加附加组件,而不必添加其它附加组件。例如,可以添加连接器而不必添加动作引擎的附加实例。在一些实施方式中,机器人系统120的一个或多个组件或组件的一部分可以在虚拟机上运行。通过在虚拟机上运行,可以根据期望发起附加的虚拟机。

[0068] 如上所述,部分地由于自然语言的微妙性和歧义性,以及输入空间(例如,可能的最终用户话语)的维度和输出空间的大小(意图的数量),构建诸如可以基于最终用户的话语来确定最终用户的意图的用户意图分类引擎之类的机器人系统是一项具有挑战性的任务。由此,可能需要监视、调试和修改新的机器人系统,以提高机器人系统的性能和机器人系统的用户体验。在许多情况下,可能难以更具体地识别出机器人系统性能低于期望的根本原因,并确定如何在不使用分析或优化工具的情况下改善机器人系统。

[0069] 在一些情况下,机器人的所有者、开发人员或管理员可能希望监视机器人的操作状态并理解机器人的使用方式以及最终用户在哪里放弃机器人,以便改进机器人。例如,机器人的所有者、开发人员或管理员可能想知道哪些机器人对话是成功的,哪些对话不成功,以便识别和诊断机器人系统表现不佳的元素。

[0070] 根据一些实施例,分析系统可以与机器人系统集成。分析系统可以监视最终用户与机器人系统之间的对话期间发生的事件、汇总和分析收集到的事件,并在图形用户界面上以不同的概括级别(诸如所有对话、不同类别的对话,以及个别对话)以图形方式提供关于对话的信息。例如,图形用户界面可以显示用于过滤或选择某些类型的对话或个别对话的选项,并且例如通过使对话的路径可视化以图形方式显示所选择的信息。分析系统还可以提供建议、选项或其它信息,以改进机器人系统。

[0071] 图2描绘了根据某些实施例的包括机器人系统(诸如,机器人系统120)和用于监视、分析、可视化和改进机器人系统的性能的机器人分析系统的集成系统200。如图所示,机器人系统可以包括连接器230和多个机器人引擎,诸如对话引擎212、意图建模器214、实体解析器216和自定义组件218。机器人系统还可以包括数据库240、管理API 250、用户界面254和UI服务器252。机器人分析系统可以包括收集器255、扩充(enrichment)引擎260、数据库270和REST服务器280。机器人分析系统还可以包括用户界面292和UI服务器290。机器人分析系统的收集器255可以收集在机器人系统处发生的事件205。可以通过用户界面292和用户界面254将来自机器人分析系统的反馈294提供给机器人系统。

[0072] 连接器230可以通过一个或多个通道(诸如,通道220和222)充当机器人系统与一个或多个最终用户之间的接口。每个通道可以是消息传递应用,诸如消息传递通道(诸如,

Facebook Messenger、Facebook WhatsApp、微信、Line、Kik、Telegram、Talk、Skype、Slack或SMS)、虚拟私人助理(诸如,Amazon Dot、Echo或Show、Google Home、Apple HomePod等)、扩展为具有聊天能力的本地或混合/响应移动应用或web应用的移动和web应用扩展,或基于语音的输入(诸如,具有使用Siri、Cortana、Google Voice或其它语音输入进行交互的界面的设备或应用)。在一些实施例中,连接器130可以标准化来自不同通道的内容,使得机器人系统可以跨不同的消息传递应用系统分析内容。内容标准化处理可以包括将内容从每种类型的消息传递应用格式化为用于处理的公共格式。在一些实施例中,机器人系统可以包括用于每个通道的一个或多个连接器。

[0073] 意图建模器214可以用于确定与最终用户话语相关联的最终用户意图。在一些实施例中,用于基于机器人系统从最终用户接收到的一条或多条消息来确定最终用户的意图的意图建模器214可以使用自然语言处理器来标记词性(动词、名词、形容词)、找到词根/词干(runs/running/ran->run)、并标记实体(德克萨斯->位置(Texas->LOCATION))。在一些实施例中,意图建模器214可以标准化消息。例如,“玛丽跑到德克萨斯(Mary ran to Texas)”可以变成“人跑到位置(PERSON run to LOCATION)”。意图建模器还可以包括用于检测最终用户消息内具有相同含义的词的逻辑。例如,如果训练数据集包括:“玛丽跑到德克萨斯(Mary ran to Texas)”和“鲍勃走到底特律(Bob walked to Detroit)”,两者都映射到相同的意图,并且走(run)/跑(walk)出现在相同的意图集中,那么意图建模器214可以了解到出于意图解析的目的走(run)=跑(walk)。在一个说明性示例中,“玛丽跑到德克萨斯(Mary ran to Texas)”可以变成“人跑到位置(PERSON run to LOCATION)”,并且“鲍勃走到底特律(Bob walked to Detroit)”可以变成“人走到位置(PERSON walk to LOCATION)”。在说明性示例中,两个句子都可以与相同的意图相关联,因为出于意图解析的目的,“名词跑到名词(noun run to noun)”与“名词走到名词(noun walk to noun)”相同。在另一个示例中,“我想要订一个大的奶酪披萨(I want to order a large cheese pizza)”和“我想要订购一个小的意大利辣肠披萨(I want to order a small pepperoni pizza)”两者可以被标准化为“我想要订购一个Bots_披萨尺寸(Bots_PizzaSize)Bots_配料(Bots_Toppings)的披萨”。

[0074] 在标准化之后,可以确定词的出现可以表示某种意图的概率。在一些示例中,可以使用基本概率算法来组合概率,就好像它们是独立的一样。例如,如果“订购(order)”表明订购比萨的概率为20%,并且“比萨(pizza)”表明订购比萨的概率为10%,那么总概率将为1-(1-0.2)(1-0.1)=28%。一些概率可能基于词的存在,或基于某些语言元素,例如否定词或人称代词的存在。

[0075] 另一个级别的规则可以是模板规则,它是词的组合。在一些示例中,训练数据集中的每个句子一旦被标准化,就可以自动成为规则。在这样的示例中,训练数据集可以包括非常少量的短句子。模板规则可以返回1的概率。可以经由归纳处理从规则中生成新规则。例如,以下句子可能属于跟踪支出:“我上个月花了多少汽油费(How much did I spend last month on gas)?”和“我5月花了多少食品费(How much did I spend in May on food)?”。这些句子可以用来推导规则“我花了多少(How much did I spend)”,因为这是它们之间共享的部分。在其它示例中,训练数据集可以包括短语“我花了多少(How much did I spend)”以达到相同的结果。

[0076] 上面描述的示例允许意图的定义不包括重复项,诸如命名实体的变体(例如,“给苏钱(Send money to Sue)”和“给鲍勃钱(Send money to Bob)”)。但是,其中一个或两个词不同的相似句子可以用于训练。类似的句子可以使模型学习哪些词对于意图解析可能具有相同含义,以及哪些词可能是常见的拼写错误。

[0077] 如果特定词或一组词(例如,动词)对意图是重要的,那么可以通过具有使用此类词(及其同义词)的更多示例以及用于其它意图的此类词的更少示例来操纵概率。

[0078] 还可以提供示例以防止模型做出错误的断言。例如,特定的子短语或仅对于某种意图出现的词可能会导致错误的断言。类似地,可以防止模型使用属于用于训练的不同意图的相似句子来合成广泛规则。

[0079] 实体解析器216可以识别与最终用户意图相关联的实体(诸如,对象)。例如,除了由意图建模器214识别的最终用户意图(诸如“订购比萨”)之外,实体解析器216还可以解析与意图相关联的实体,诸如比萨类型、比萨配料(toppings)等。

[0080] 对话引擎212可以用于处理最终用户与机器人系统之间的对话。例如,对话引擎212可以基于由意图建模器214识别的最终用户意图以及与由实体解析器216识别的与最终用户意图相关联的实体来响应最终用户话语。在一些实施例中,对话引擎212可以使用状态机以应对与最终用户的对话,该状态机包括用户定义的状态(诸如,最终用户意图)和在状态中或从一个状态到另一个状态要采取的动作。

[0081] 自定义组件218可以包括用于特定机器人系统的定制模块。例如,金融机器人可以包括可以用于例如检查余额、转移资金或支付账单的自定义组件。

[0082] 数据库240可以用于存储机器人系统的数据,诸如用于分类模型的数据、对话日志等。机器人系统的管理员或开发人员可以使用管理API 250来管理机器人系统,诸如重新训练分类模型、编辑意图或以其它方式修改机器人系统。管理员或开发人员可以使用用户界面254和UI服务器252来管理机器人系统。

[0083] 当机器人系统正在运行时,可能会生成各种事件。可以基于机器人系统中包括的一条或多条指令来生成事件。例如,当机器人系统进入特定状态时,可以生成事件,其中特定状态由机器人系统的管理员或开发人员定义。随着事件的生成,机器人分析系统可以收集、存储和分析事件。当捕获事件时,还可以收集与事件相关联的附加信息,其中附加信息可以指示其中生成事件的当前上下文。

[0084] 例如,对话事件可以由对话引擎212生成。对话事件可以包括机器人系统从最终用户设备接收到的消息(称为msg_received)。Msg_received可以包括以下参数或变量中的一个或多个:消息的内容、机器人系统接收到消息时的时间、接收到的消息的语言、设备特性(例如,版本或名称)、操作系统特性(例如,版本或名称)、地理位置特性(例如,互联网协议地址、纬度、经度等)、标识信息(例如,用户ID、会话ID、机器人系统ID、租户ID等)、时间戳(例如,设备被创建、设备被发送、收集器被派生的时间戳)、通道等。

[0085] 对话事件还可以包括由机器人系统发送给最终用户设备的消息(称为msg_sent)。Msg_sent可以包括以下中的一项或多项:消息的内容(例如,消息的文本或HTML)、机器人系统发送消息时的时间、消息的语言、消息的创建者(例如,机器人系统或最终用户设备)、设备特性、操作系统特性、浏览器特性(例如,版本或名称)、应用特性(例如,版本或名称)、地理位置特性(例如,互联网协议地址、纬度、经度等)、标识信息(例如,用户ID、会话ID、机器

人系统ID、租户ID等)、通道(例如,Facebook或Webhook)等。

[0086] 对话引擎212还可以生成对话状态执行事件。如上所述,对话引擎212可以使用状态机来确定与最终用户的对话流。状态机可以包括一组状态和状态之间的转换规则。对话引擎212可以为每个最终用户对话执行状态机,并且可以为对话引擎212逐步通过以处理最终用户话语的每个状态产生对话状态执行事件。对话状态执行事件的属性可以包括例如状态名称、组件名称、下一个动作、实体匹配、意图匹配、变量、用户查询语句、响应语句、执行时间、通信语言、设备特性、操作系统特性、地理位置特性、标识信息、时间戳、通道等。状态名称可以是当前执行的状态的名称或者是“错误状态”。组件名称可以是针对当前状态执行的机器人组件的名称。下一个动作可以是要执行的下一个动作。实体匹配可以是在当前消息中解析的实体。意图匹配可以是用分数值解析的意图。变量可以是当前状态的变量值。查询语句可以是最终用户发送的消息。响应语句可以是发送给最终用户的消息。执行时间可以是完成状态执行的时间戳。通信语言可以是被交谈的消息的语言。设备和/或操作系统特性可以与与机器人系统交互的最终用户相关联。浏览器和/或应用特性可以与与机器人系统交互的最终用户相关联。地理位置特性可以是与机器人系统交互的最终用户的位置。

[0087] 由于意图建模器214的执行,可能发生意图解析事件。意图建模器214可以使用经训练的或其它方式定义的分类模型来基于最终用户话语从一组意图中识别最终用户意图。意图分类的结果可以被捕获为意图解析事件属性,其可以包括例如最终意图分类结果(诸如,识别出的意图)和与意图集中的每个相应意图相关联的置信度分数。

[0088] 实体解析器216可以生成实体解析器事件。实体是与最终用户意图关联的对象。当创建机器人系统时,可以确定实体定义规则。例如,除了解析诸如“订购比萨”之类的最终用户意图之外,机器人系统还可以使用实体解析器216来解析相关联的实体,诸如比萨类型、比萨配料等。可以在实体解析时捕获实体解析器事件。与实体解析器事件相关联的属性的示例可以包括实体名称、所应用的规则、搜索项、已解析的状态、查询语句、实体类型、执行时间、通信语言、设备特性、操作系统特性、浏览器特性、应用特性、地理位置特性、标识信息、时间戳、通道等。实体名称可以是当前被解析的实体的名称。所应用的规则可以是例如在前、在后或汇总。搜索项可以是从(from)、到(to)、目的地(destination)、起源(origin)等。解析的状态可以是针对实体解析的对话状态。查询语句可以是包含实体值的消息。实体类型可以是系统或派生的。执行时间可以是实体解析的时间戳。通信语言可以是交谈的消息的语言。设备和/或操作系统特性可以与和机器人系统交互的最终用户相关联。浏览器和/或应用特性可以与和机器人系统交互的最终用户相关联。地理位置特性可以是和机器人系统交互的最终用户的位置。

[0089] 自定义组件也可以生成事件,诸如预定义事件或自定义事件。预定义事件可以是在执行自定义组件时捕获的特性。预定义事件的属性的示例可以包括:组件名称、事件名称、有效载荷、执行时间、通信语言、设备特性、操作系统特性、浏览器特性、应用特性、地理位置特性、标识信息、时间戳、通道等。组件名称可以是当前被执行的自定义组件的名称。事件名称可以是invoked(被调用)、invocation_failed(调用失败)、replied(回复)、replied_failed(回复失败)等。在失败的情况下,有效载荷可以是失败的原因、堆栈跟踪等。执行时间可以是指示事件何时发生的时间戳。通信语言可以是交谈的消息的语言。设备和/或操作系统特性可以与和机器人系统交互的最终用户相关联。浏览器和/或应用特性可

以与和机器人系统交互的最终用户相关联。地理位置特性可以是与机器人系统交互的最终用户的位置。

[0090] 自定义组件还可以在自定义组件的执行期间发出自定义事件。自定义事件的属性的示例可以包括组件名称、事件名称、自定义有效载荷、执行时间、通信语言、设备特性、操作系统特性、浏览器特性、应用特性、地理位置特性、标识信息、时间戳、通道等。组件名称可以是当前被执行的自定义组件的名称。事件名称可以是用户定义的事件名称(例如,Balance_Retrieved(余额查询))。有效载荷可以是例如{“amount(账户)”:“USD 100(100美金)”,“account(账户)”:“checking(核算)”}。执行时间可以是指示事件何时发生的时间戳。通信语言可以是交谈的消息的语言。设备和/或操作系统特性可以与和机器人系统交互的最终用户相关联。浏览器和/或应用特性可以与和机器人系统交互的最终用户相关联。地理位置特性可以是和机器人系统交互的最终用户的位置。

[0091] 错误事件和超时事件也可以由机器人系统在执行期间生成。当发生错误时,可以生成错误事件。当最终用户对话在一段时间内处于非活动状态时,可以生成超时事件,该时间段可以在通道处进行配置。

[0092] 机器人分析系统可以在机器人系统与最终用户进行对话并生成对应事件时收集事件和附加信息。例如,收集器255可以收集事件和附加信息,并将收集的信息发送到队列。在一些实施例中,收集器255可以是可配置的并且可以被编程为根据需要收集上述的不同事件和/或事件属性。例如,收集器255可以被配置为捕获对话状态属性、意图解析属性、实体解析属性以及错误和超时属性。在一些实施例中,收集器255还可以被配置为收集关于由除机器人系统以外的系统生成的事件295的信息。

[0093] 扩充引擎260可以对收集到的事件和其它信息执行验证和扩充,并将它们写入到数据库270。例如,基于收集到的IP地址、扩充引擎260可以确定与IP地址相关联的最终用户的位置。作为另一个示例,扩充引擎260可以从收集到的信息中提取某些特征,诸如确定最终用户使用的web浏览器或通道。REST服务器280可以分析扩充的事件和其它信息,并基于某些汇总度量272生成各种报告。可以通过UI服务器290在用户界面292上将报告显示给机器人系统的所有者、管理员或开发人员。机器人系统的所有者、管理员或开发人员可以向机器人系统提供反馈294,以改进机器人系统。

[0094] 如上所述,构建可以基于用户话语确定最终用户的意图的机器人系统,诸如用户意图分类引擎(例如,消息处理器150或意图建模器214),是一项具有挑战性的任务,部分原因是自然语言的微妙性和歧义性、输入空间(例如,可能的用户话语)的维度和输出空间的大小(意图的数量)。在一些机器人系统中,训练话语和相关联的意图可以用于训练机器人以基于用户话语识别意图。例如,在用于鞋店的可能关注捕获用户“购买鞋子”意图的机器人系统中,“购买鞋子”的高级别意图可以馈送到用户意图分类引擎中,并且也可以提供诸如“我想要购买一些鞋子(I want to buy some shoes)”、“我想找一双鞋子(I am looking for a pair of shoes)”、“我想要鞋子(Iwant shoes)”等训练短语来训练用户意图分类引擎。在一些实施例中,用户意图分类引擎可以尝试扩展示例短语并使用示例短语来匹配用户话语。例如,用户意图分类引擎可以能够将短语“我想要购买靴子(I want to purchases boots)”与意图“购买鞋子(purchase shoes)”进行匹配,因为自然语言处理器(例如,自然语言处理器152)可以理解靴子、鞋子或者甚至拖鞋之间的语义距离很小并且它们都指相同

类别的事物。此类用户意图分类引擎可能对于一些聊天机器人交互适用,但可能无法应对更复杂的任务。

[0095] 代替如上所述使用意图匹配,一些机器人系统可以将较低级别的分析结果(语法分析、实体、关键短语)与特定领域的知识图相组合,以识别用户意图。知识图(或本体)可以将意图、短语、名词和形容词与领域中的概念相连。例如,机器人可以使用自然语言处理器来分析来自用户的输入,并使用分析结果来浏览知识图并在语义级别上找到用户意图。然后,用户意图可以用于确定对用户的最适当的下一个响应。

[0096] 知识图获取信息并将信息集成到本体中,该本体包括主语范围或领域中的一组概念和类别,并显示其特性以及它们之间的关系。在一些情况下,知识图还可以应用推理器来推导新知识。知识图可以捕获关于地点、人、事物、处理、应用以及它们之间的关系的事实,并使用此类信息向执行搜索或查询的用户提供例如更多相关信息。知识图通常使用有向图来表示知识库,该有向图在节点或顶点(表示实体)之间具有带标记的边(表示关系)。例如,知识图可以基于一组(主语,谓语,宾语)形式的三元组,诸如(奥巴马(Obama),出生于(bornin),夏威夷(Hawaii))、(巴西足球队(Brazilian football team),有位置(has position),守门员(goalkeeper))等。

[0097] 图3图示了知识图300的示例。知识图300可以包括多个顶点310,其表示例如地点(诸如,巴黎(Paris))、人(诸如,Bob、Bill和莱昂纳多·达·芬奇(Leonardo da Vinci))、事物(诸如,埃菲尔铁塔(Eiffel Tower)和蒙娜丽莎(Mona Lisa))、概念(例如,时间、年龄、维度和事物的类别,诸如城市、人和地方)、处理、应用等。顶点310通过带标记的边320连接,该带标记的边320描述了每对连接的顶点中的两个顶点之间的关系。边320可以是单向或双向的。通过沿着由边320形成的路径遍历知识图300,一个顶点310可以从另一个顶点310被检索到。

[0098] 图4A-4D图示了基于诸如书或维基百科之类的知识库来构建知识图的处理的示例。在图4A中,可以从知识库获得串410(例如,“莱昂纳多·达·芬奇(Leonardo da Vinci)”)。在图4B中,可以从串410中识别和辨识实体420(例如,“莱昂纳多·达·芬奇(Leonardo da Vinci)”)。在图4C中,可以从知识库中识别与实体420相关的实体432,并且可以将其添加到图谱(graph)430。在图4D中,可以从知识库确定实体432之间的关系,并将其作为实体432之间的边442或连接添加到图上,以形成包括多个实体432和边442的知识图440。实体432和边442可以由(主语,谓语,宾语)、(实体,属性,值)或其它形式的三元组表示。边442上的标签可以指示实体432之间的关系。

[0099] 知识图440可以用作种子图来构建更大规模的知识图。例如,可以从知识库或参考知识库(例如,外部知识库)中识别或确定知识图440中的实体432的相关实体以及知识图440中的相关实体与实体432之间的关系,并将其添加到知识图440以进一步扩展知识图440。知识图440的扩展可以继续,直到例如穷尽给定领域中的知识为止,这可以产生给定领域的知识图。

[0100] 知识图在自然语言理解中的应用包括QnA,诸如通过嵌入知识图的相关部分来回答问题,或者使用深度卷积神经网络学习的奖励函数来生成相关查询子图。知识图还可以例如由搜索引擎用来响应于用户搜索而提供更多相关信息,或者由聊天机器人用来执行如上所述的意图分类。

[0101] 聊天机器人在具有足够数据的特定领域中可能表现更好。对于给定的机器人,用户提供的数据集可能仅包括例如大约50-100个话语。这个有限的数据集可能不足以构建可以被机器人用来正确工作的知识图。参考知识图或知识库,诸如WikiData、Freebase、DBpedia等,可以用于扩展特定知识领域中的知识并生成定制的知识图。

[0102] 在一些实施例中,可以通过例如随机地游走参考知识图或其它知识库,并且基于例如已访问参考知识图或知识库的节点的频率从参考知识图或知识库中选择信息来扩展知识图。但是,对于诸如聊天机器人之类的应用而言,通过随机游走生成的知识图的质量可能不足。例如,通过随机游走生成的知识图可能有噪声。

[0103] 根据某些实施例,公开了使用客户数据集(包括例如用户话语、意图、实体以及问题和答案(QnA))和参考(例如,外部)知识库(诸如,WikiData)为给定领域(例如,特定机器人)构建定制知识图的方法。定制的知识图可以是特定于领域的,并且可以在两个主要阶段中生成。首先,可以从客户数据集中提取实体,诸如用户提供的话语,并将其用作初始知识图的顶点或节点,该初始知识图可以用作扩展的种子图。这些提取出的实体及其之间的链接可以用于形成种子图,以初始化定制的知识图。在第二阶段,可以通过使用有限状态机遍历参考知识图来扩展种子图,其中可以将种子图中的实体映射到参考知识图作为遍历参考知识图并选择要添加到定制的知识图的候选实体的起始顶点。从参考知识图中选择的候选实体然后可以基于添加到定制知识图的其优先级分数进行过滤。

[0104] 图5是图示根据某些实施例的用于构建定制的知识图(诸如,用于聊天机器人)的方法的示例的简化流程图500。图5中呈现的和下面描述的处理旨在是说明性和非限制性的。虽然图5描绘了以特定顺序或次序发生的各种操作,但这并不旨在限制。在某些替代实施例中,可以以一些不同的次序执行操作,或者也可以并行执行一些操作。此外,该方法的实施例可以用硬件、软件、固件、中间件、微代码、硬件描述语言或其任意组合来实现。当用软件、固件、中间件或微代码来实现时,用于执行相关联的任务的程序代码或代码段可以存储在诸如存储介质的计算机可读介质中。一个或多个处理器可以执行相关联的任务。

[0105] 在502处,计算系统可以接收用户数据集。用户数据集可以包括例如用户话语、用户定义的实体和意图、用户提供的问题和答案等。如上所述,在一些实施例中,用户数据集可以包括例如大约50至大约100个话语,其中每个话语可以或可以不用相关联的意图标记。

[0106] 在504处,可以使用技术从诸如用户话语的用户数据集中提取实体,这些技术诸如词性(PoS)标记(参见例如,Kristina Toutanova等人,“Feature-Rich Part-of-Speech Tagging with aCyclic Dependency Network”,Proceedings of HLT-NAACL 2003,第252-259页)、命名实体辨识(NER)(参见例如,Jenny Rose Finkel等人,“Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling”,Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics(ACL 2005),第563-370页)、使用例如CoreNLP进行构成解析(参见例如,Manning等人,“The Stanford CoreNLP Natural Language Processing Toolkit”,Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics:System Demonstrations,2014,第55-60页)等。在一些实施例中,可以使用诸如Dexter 2技术之类的实体链接技术从用户数据集中检测实体(参见例如,Ceccarelli等人,“Dexter:an open source framework for entity linking”,Proceedings of the

sixth international workshop on Exploiting semantic annotations in information retrieval,ACM 2013,第17-20页)。

[0107] 在506处,可以基于用户数据集例如使用上面描述的用于提取实体的技术来识别识别出的实体之间的关系。该关系可以用于链接或连接识别出的实体,以生成种子图,该种子图包括识别出的实体(作为种子图中的节点或顶点)以及识别出的实体之间的关系(作为顶点之间的边)。边可以是定向的边,包括实体之间的边的方向。边可以是单向或双向的。种子图可以包括少量顶点,并且可以是可能不包括实体之间的所有关系的稀疏图。种子图可以包括一个或多个弱连接的组件(WCC),其中WCC通常是指有向图的最大子图,使得可以通过沿着无向路径(即,忽略边的方向)遍历边来从任何其它节点开始到达任何节点。弱连接的组件中的节点都可以具有至少1的出度(out-degree)或入度(in-degree)。作为对照,强连接的组件可以指以下这种子图,其中子图中存在从每个节点到每个其它节点的有向路径。

[0108] 在508-516处,可以使用模拟退火优化方法(参见例如,Nourani&Andresen,“A comparison of simulated annealing cooling strategies”,J.Phys.A:Math.Gen.51.1998,8373-8385)通过逐步向种子图添加实体和关系来迭代地扩展种子图。在每次迭代中,可以在508处设置或调节温度值。如下面详细描述的,该温度可以用于确定要添加到知识图的候选实体的优先级分数。

[0109] 在510处,对于种子图中的每个弱连接的组件,可以将弱连接的组件中的每个实体(即,顶点)映射到参考知识图,并且将其用作用于遍历参考知识图和使用如下面更详细描述的有限状态机识别多达M个候选实体的起始顶点。识别出的候选实体的资源描述框架(RDF)可以被保存为优先级队列中的条目。

[0110] 在512处,可以使用作为温度的函数的评分函数对优先级队列中的条目进行评分。在一些实施例中,在510处识别并添加到优先级队列的候选实体的总数可以大于某个阈值数N。可以执行诸如蓄水池抽样技术(参见例如,J.Vitter,“Random sampling with areservoir”,ACM Trans.on Mathematical Software 11(1):37-57,1985年3月)的技术以将优先级队列中的条目的数量减少为N,从而减少用于计算条目的优先级分数的时间,计算条目的优先级分数可能比知识图遍历在计算上更加昂贵。可以为N个条目中的N个候选实体中的每一个以及到达候选实体的对应路径计算优先级分数,如下文更详细描述的。

[0111] 在514处,可以从优先级队列中具有最高分数的多达N个条目中选择K个条目,并且可以将由K个条目识别的对应实体和关系添加到种子图以扩展种子图。

[0112] 在516处,如果确定温度高于阈值,那么可以在下一次迭代中再次执行508-514处的操作,在下一次迭代期间可以改变温度(并因此改变评分函数),并且弱连接的组件中的实体可能不同,因为已使用优先级队列中的前K个条目对种子图进行了修改。当温度低于阈值时,定制的知识图的优化可以结束,并且定制的知识图可以用于与用户数据集相关联的用户应用,诸如特定的机器人或搜索引擎。

[0113] 表1示出了用于实现生成定制知识图的方法的伪代码的示例,该方法诸如以上关于图5所描述的方法。表1中的伪代码示出了用于遍历参考知识图的有限状态机的示例。在表1中,伪代码的第一行涉及基于用户数据集(例如,用户话语)识别实体和实体之间的链接。伪代码的第2行涉及创建种子图。第3行用于设置模拟退火处理的初始条件,以扩展种子

图来生成定制的知识图,包括初始化温度和优先级队列。第4-23行涉及在模拟退火处理的每个温度下执行的操作,其中第5-15行描述了使用有限状态机遍历参考知识图以识别候选实体并将候选实体的RDF三元组保存为优先级队列中的条目,该优先级队列将在下面更详细描述。第16-17行涉及在优先级队列中的条目的数量大于阈值数量N时减少优先级队列中的条目的数量以进行优先级分数计算。第18-22行涉及计算优先级队列中的条目的优先级分数,并使用优先级分数最高的前K个条目来扩展种子图。在第23行处更新温度。

[0114] 表1知识图生成

[0115]

```
1: V ← 实体集合，E ← 任何两个 v ∈V 之间的所有真链接
2:  G ← (V，E)
3:  T ← T0 ，pq ← 空优先级队列
4:  while T > Tmin do
5:    for wg ∈弱连接的组件(G) do
6:      fsm ← 根据图 4 的有限状态机，cache ← φ
7:      while State(fsm) ≠ end do
8:        for w ∈wg do
9:          if w ∈消歧顶点 then
10:           移动 fsm 到消歧状态
11:         else if w ∈禁止顶点 then
12:           移动 fsm 到禁止状态
13:         else
14:           移动 fsm 到实体状态
15:           cache ← cache + w
16:   if length(cache) > N then
17:     cache ← 蓄水池抽样(cache)
18:   for v ∈cache do
19:     优先级三元组← 计算优先级(v)
20:     pq ← pq + 优先级三元组
21:   G' = pq 中的前 K 个 RDF 三元组
22:   G ← G ∪G'
23:   T ← 更新温度(T )
```

[0116] 以上相对于图3和表1所述的方法通常可以包括:(1)从用户提供的话语中提取实体,并基于提取出的实体和实体之间的链接构建种子图,以及(2)通过使用有限状态机重复遍历参考知识图来扩展种子图,下面将对其中每一个进行更详细的描述。

[0117] 从诸如一组原始句子(诸如,话语)之类的用户数据集中,可以提取实体并将其用作知识图顶点,并且可以确定实体之间的现有关系并将其用作知识图中的顶点之间的边。例如,可以对话语执行以上关于图5的504描述的词性标记、命名实体辨识和使用CoreNLP技术的构成解析来生成候选实体列表。PoS标记技术包括用词的适当词性(诸如名词、动词、形容词、副词、代词等)标记词。命名实体辨识(也称为实体识别、实体组块或实体提取)包括将非结构化文本中的命名实体定位和分类为预定义类别,诸如名称、组织、位置、医疗代码、时间表达、数量、货币值、百分比等。构成分析包括从句子中提取基于构成的解析树,以根据短语结构语法表示句子的句法结构。例如,句子“John hit the ball(约翰击中球)”的基于构成的解析树可以从句子(S)开始,并在每个叶子节点(John,hit,the,ball)中结束,其中“John”可以是名词(N)节点,“hit”可以是动词(V)节点(在这个示例中为及物动词),“the ball”可以是作为分支节点的名词短语(NP),“the”可以是限定词(D或DT)节点,并且“ball”

可以是名词(N)节点。

[0118] 在一些实施例中,当生成候选实体列表时,可以应用以下预处理规则:(a)检查NER是否不为空以及词是否是名词短语(例如,

[0119] “apple tree(苹果树)”或“the ball(球))”)的一部分并由NER进行标记;(b)检查其中专有名词(NNP,其命名特定的人、地点、事物或想法,例如,“John”)由NP主导(即,NP在语义解析树中是NNP的父级)的匹配项(提取出的实体);(c)检查其中普通名词(NN,其是命名不特定的人、地点、事物或想法的通用名词,例如,“desk(桌子)”)由NP主导(即,NP在语义分析树中是NN的父级)的匹配项(提取出的实体);(d)从上述剥离人称代词(PRP,例如,“I(我)”,“he(他)”或“she(她)”)和限定词(DT);(e)将NER短语划分成组块;以及(f)从NER短语中移除数字和日期。例如,将以上规则应用于句子“How much did I spend from my checking account on ServiceA(我在服务A上从我的支票账户花了多少钱)?”可以产生包括['checking account(支票账户)','ServiceA(服务A)']的候选实体的列表。在另一个示例中,将以上规则应用于句子“I want to go to an outdoor cafe on Pioneer Day(我想要在Pioneer Day去户外咖啡店)”可以产生包括['Pioneer Day','outdoor cafe(户外咖啡店)','outdoor cafe on Pioneer Day']的候选实体的列表。

[0120] 在一些实施例中,上面讨论的Dexter 2技术还可以用于通过检测并行引用实体的文本片段来生成候选实体的列表。然后可以将候选实体的两个列表组合来生成候选实体的列表。然后可以使用候选实体的组合列表来生成用于知识图生成的种子图。

[0121] 在一些实施例中,Dexter 2技术还可以用于将候选实体链接到诸如WikiData之类的参考知识图中的其对应条目,其中Dexter 2技术可以用于将候选实体链接到例如Wikipedia条目,然后可以将Wikipedia条目映射到WikiData条目。例如,从示例句子“How much did I spend from my checking account on ServiceA?”中提取出的实体(及其对应的WikiData ID)可以包括“ServiceA”(Q780442)和“checking account”(Q695430)。在另一个示例中,从示例句子“Michael Jordan played basketball at an outdoor cafe on Pioneer Day(迈克尔·乔丹在Pioneer Day在户外咖啡店打篮球)”中提取出的实体(及其WikiData ID)可以包括“café”(Q30022)、“Pioneer Day”(Q254295)、“basketball”(Q5372)、“Michael Jordan”(Q41421)和“outdoor”(Q3312129)。在一些实施例中,Mediawiki Webservice API可以用于搜索与上述方法不匹配的实体。

[0122] 表2图示了用于从用户数据集(例如,话语)中提取实体并构建种子图的方法(诸如上述方法)的示例。在表2中所示的操作1中,例如,使用词性标记、命名实体辨识或使用如上所述的CoreNLP技术的构成解析从用户话语的原始句子中提取实体。

[0123] 表2提取实体和构建种子图的方法

1. 从原始句子中提取实体：
    a. 词性标记(PoS)
    b. 命名实体辨识 (NER)
    c. 使用 CoreNLP 的构成解析
2. 应用预处理规则：
    a. 检查命名实体辨识（NER）是否不为空以及词是否为名词短语（NP）的一部分
    b. 其中由 NP 主导的专有名词（NNP）的匹配项
    c. 由 NP 主导的普通名词（NN）的匹配项
    d. 上述的剥离代词（PRP）和限定词（DT）

[0124]

    e. 将 NER 短语划分为组块
    f. 从 NER 短语中移除数字
3. 基于规则，根据解析结果和通过实体链接技术（例如，Dexter 2.0）生成的实体候选列表确定图形查找查询
4. 对于通过 Dexter 2.0 不匹配的实体，对 Mediawiki Webservice API 进行调用
5. 使用 Redis 高速缓存返回的 WikiData 实体 ID，以减少将来的 API 调用
6. 对于话语中的所有实体，检查所有可能的组合（例如，N 选 2 组合）
    a. 如果话语中任意两个实体之间存在链接，那么将链接添加到图 G
        * 也可以执行上述步骤来检查整个数据集中所有实体之间的链接

[0125] 在操作2中,如上所述,可以应用一组预处理规则来生成候选实体列表。预处理规则可以包括例如(a)检查命名实体辨识是否不为空以及词是否为名词短语的一部分;(b)检查其中专有名词由NP主导(即,NP在语义解析树中是NNP的父级)的匹配项(提取出的实体);(c)检查其中普通名词由NP主导(即,NP在语义解析树中是NN的父级)的匹配项(提取出的实体);(d)从上述剥离代词(PRP)和限定词(DT);(e)将NER短语划分成组块;以及(f)从NER短语中移除数字和日期。上述操作之后的最终列表是每个单独规则的一组结果。

[0126] 在操作3中,可以基于上述预处理规则,根据操作1中的话语解析的结果以及通过实体链接技术(诸如,上述Dexter 2.0框架)生成的候选实体列表,来确定图形查找查询。如本文所使用的,实体链接通常可以指在存在歧义时将实体与正确的知识库条目进行匹配。作为示例,从句子“On July 20,1969,the Apollo 11astronauts Neil Armstrong,Michael Collins,and Edwin 'Buzz' Aldrin Jr.realized President Kennedy's dream(1969年7月20日,阿波罗11号宇航员Neil Armstrong、Michael Collins和Edwin 'Buzz' Aldrin Jr.实现了肯尼迪总统的梦想)”中,可以确定以下图形查找查询:“Apollo11”(航天任务),“astronauts”,“Neil Armstrong”(宇航员),“Michael Collins”(宇航员),“Edwin 'Buzz' Aldrin”(宇航员)和John F.Kennedy”。在另一个示例中,从句子“Michael Jordan played for the Chicago Bulls(迈克尔·乔丹在芝加哥公牛队打球)”中,可以确定以下图形查找查询:“Michael Jordan”(篮球运动员)和“Chicago Bulls”。Dexter 2技术还可以用于将图形查找查询链接到诸如参考知识图(例如,WikiData)之类的参考知识库中其对应的条目。例如,可以使用Dexter 2技术将图查询查询链接到Wikipedia条目,然后可以将Wikipedia条目映射到WikiData条目。

[0127] 在操作4中,对于使用例如Dexter 2.0不与参考知识库条目匹配的每个实体,可以

对Mediawiki Webservice API进行调用,以进行实体链接和/或从参考知识库获取信息。在一些实施例中,在操作5中,可以例如使用Redis将返回的WikiData实体ID进行高速缓存以减少将来的API调用。

[0128] 在操作6中,对于每个话语中的所有实体,可以检查所有可能的组合(N选2组合)。如果话语中两个实体之间存在链接,那么可以将该链接添加到种子图。在操作6的结束,可以生成种子图G=(V,E),其中顶点V包括识别出的实体的集合,并且边E包括在实体中的任何两个实体之间的链接。还可以对整个用户数据集中的实体执行两个实体之间的链接检查。

[0129] 如上所述,用户数据集可以小,诸如大约50-100个话语。因此,种子图可能仅包括少量实体,诸如大约50-100个实体。另外,种子图中的实体之间的链接可能是稀疏的。种子图可能需要扩展以包括对给定领域(诸如,特定聊天机器人)有用的附加信息。根据某些实施例,可以使用有限状态机来遍历诸如WikiData之类的参考知识图,以识别可以添加到种子图以生成定制知识图的附加信息(例如,实体和链接)。在一些实施例中,优先级队列和优先级函数可以用于确定要添加到知识图的实体和/或链接。在一些实施例中,可以使用模拟退火处理来迭代地优化定制知识图。表3中所示的伪代码描述了一种示例方法。

[0130] 表3.扩展种子图

[0131]

1. 初始化 $T \leftarrow T_0$，优先级队列 (pq) ← 空
2. While $T > T_{min}$
   a. for 图 G 的弱连接组件中的组件 wg（表示为 WCC（G））
      1) 初始化有限状态机(fsm)
      2) *pq_cache* ←空
      3) whilefsm.state 不是'结束'并且在*实体*状态已被访问 M 次之前对于组件 w_g 中的每个顶点 w:
         i. if 实体是消歧顶点
            移动到*消歧*状态；
            *else*if 实体是禁止顶点
            移动到*禁止*状态；
            else
            移动到*实体*状态→ 添加到 *pq_cache*
         ii. if *实体*状态已经被访问 M 次，移动到'结束'
   b. 计算 *pq_cache* 中每个实体 *e* 的分数并将其插入到 pq 中

1) 出于时间优化目的，if length(*pq_cache*) > N，则执行蓄水池抽样以将数量减少到N

2) S 是评分函数

$$S(G',t)=\frac{1}{Z}count(WCC(G'))e^{\left(\frac{-k}{T(t)}\right)}+\eta\rho(G')\left(1-\frac{1}{Z}e^{\left(\frac{-k}{T(t)}\right)}\right);k,\eta,Z\in\mathbb{R}$$

其中k，ηand Z 是实常数

3) G'由 G 与（w，e）之间的边集合的并集形成

[0132] $$G'=G\cup(w,e)$$

4) ρ(G) 是图密度:

$$\rho(G)=\frac{|E|}{|V|(|V|-1)},G=(V,E)$$

其中G是具有边E和顶点V的图，|E|是G中的边数，|V|是G中的顶点数

c. 从优先级队列中弹出前 K 个RDF三元组并将其添加到G

d. 更新温度

[0133] 在表3中的伪代码所示的方法的操作1中,可以通过设置初始温度并初始化优先级队列(诸如创建空的优先级队列)来启动模拟退火处理。

[0134] 在操作2(a)中,种子图G中每个弱连接的组件w_g中的每个实体可以映射到参考知识图,诸如WikiData,并且可以使用有限状态机来控制参考知识图从弱连接的组件w_g中的每个实体的遍历,以识别候选实体并将它们存储为优先级队列缓冲区(例如,高速缓存)中的条目。例如,如果所遍历的顶点是消歧顶点,那么状态机可以进行到消歧状态;如果所遍历的顶点是禁止顶点,那么状态机可以进行到禁止状态;并且如果所遍历的顶点是实体顶点,那么状态机可以进行到实体状态;并且如果对于每个弱连接的组件w_g,实体状态已经被访问M次,那么状态机可以进行到“结束”状态。如本文所使用的,消歧顶点是指区分共享相同或相似标签的多个条目的知识图条目或知识库条目。如本文所使用的,禁止顶点是指知识图(例如,WikiData)中增加了定制知识图的稀疏性或者没有向定制知识图提供任何有意义的信息并且因此不适合添加到定制知识图的顶点。禁止顶点可以包括例如日期和/或时间(例如,持续时间、时间点等),指向其它参考知识库标识符(IMDb、Freebase等)的链接、Wikimedia项目、文章和列表页面、主要Wikimedia知识树之外的页面以及重复和重定向的条目。

[0135] 图6图示了根据某些实施例的用于在模拟退火处理的每次迭代中控制参考知识图的遍历的有限状态机(FSM)600的示例。在FSM 600的开始状态610处,遍历可以从参考知识图中的顶点开始,其中该顶点可以对应于种子图G中弱连接组件w_g中的实体w。

[0136] 如果连接到起始顶点的下一个顶点是消歧顶点,那么FSM 600可以移动到消歧状态640。例如,每当到达WikiData中的消歧顶点时,就到达消歧状态640。如果下一个顶点不是消歧顶点,而是禁止顶点,那么FSM 600可以移动到禁止状态630。参考知识图(例如,WikiData)中不是消歧顶点或禁止顶点的任何条目都是实体w并且可以使FSM 600移动到实体状态620,其中实体w的资源描述框架(RDF)可以被添加到优先级队列缓冲区(诸如,高速缓存)。RDF可以是例如主语-谓语-宾语形式的三元组。主语可以指示资源,宾语可以指示资

源的特质或方面,并且谓语可以指示主语和宾语之间的关系。

[0137] 可以允许实体、消歧和禁止状态之间的所有可能的转变,但从禁止状态630回到禁止状态630的转变除外,因为不希望将穿过禁止顶点的任何路径添加到知识图中。通过该处理,可以识别用户数据集(和种子图)中的实体的紧密相邻实体,可以使用消歧顶点来扩展图搜索,并且可以忽略禁止顶点。当对于每个弱连接的组件w_g实体状态620已被访问M次时,或者当没有其它直接或间接链接的顶点要遍历时,可以到达结束状态650。迭代中参考知识图的遍历可以在结束状态650处停止。

[0138] 在一个示例中,如果起始顶点具有到参考知识图中的顶点“Obama(奥巴马)”的链接,那么FSM 600可以从起始状态610移动到实体状态620,并且可以将用于顶点“Obama”的RDF添加到优先级队列缓冲区(例如,pq_cache)。如果顶点“Obama”有2个链接到它的顶点:“America(美国)”(例如,通过关系“president of(是……的总统)”链接)和“20January 2009(2009年1月20日)”(例如,通过关系的“start time(起始时间)”链接),那么当遍历顶点“20January 2009”时,FSM 600可以进入禁止状态630,因为日期和时间顶点是禁止顶点。所遍历的下一个顶点是顶点“America”,它是有效实体。因此,FSM 600可以再次进入实体状态620,并将用于“America”的RDF三元组添加到优先级队列缓冲区pq_cache。该处理可以用链接到“America”和/或种子图G中弱连接组件w_g中的其它实体的顶点重复,直到实体状态620已被访问M次或没有其它直接或间接链接的顶点要遍历时为止。

[0139] 在表3的操作2(b)中,可以使用作为温度的函数的评分函数为优先级队列缓冲区中的每个条目计算优先级分数。在一些实施例中,如果在迭代期间已被访问的实体的总数超过阈值N,那么可以执行如上关于图5的512描述的蓄水池抽样技术以将优先级队列缓冲区中的实体的数量减少到N,从而减少优先级分数的计算。出于时间优化目的可以这样做,因为与使用有限状态机的知识图遍历相比,计算优先级分数可能在计算上更加昂贵。

[0140] 例如,在迭代的操作2(a)中,在识别候选实体并将对应的RDF三元组保存到优先级队列缓冲区之后,可以根据优先级函数为优先级队列中的每个实体e和到达实体e所采取的对应路径计算优先级分数:

[0141] $$S(G',t)=\frac{1}{Z}count(WCC(G'))e^{\left(\frac{-k}{T(t)}\right)}+\eta\rho(G')\left(1-\frac{1}{Z}e^{\left(\frac{-k}{T(t)}\right)}\right);k,\eta,Z\in\mathbb{R} \quad (1)$$

,

[0142] 其中WCC(G')是弱连接组件函数,用于识别图中的弱连接组件;t是当前时间步长;T(t)是时间t时的温度;k、η和Z是实常数;G'是由图G与(w,e)之间的边集的并集形成的新图。新图G'可以被表示为:

[0143] $G'=G\cup(w,e)$ (2)

[0144] ρ(G)对应于图密度并且可以根据以下等式确定:

[0145] $$\rho(G)=\frac{|E|}{|V|(|V|-1)},G=(V,E) \quad (3)$$

,

[0146] 其中G是具有边E和顶点V的图,|E|是图G中的边数,并且|V|是图G中的顶点数。

[0147] 在表3的操作2(c)中,可以选择优先级队列中具有最高优先级分数的条目并将其添加到种子图G中。在操作2(d)中,可以更新模拟退火处理的温度以发起操作2中处理的新迭代。时间步长t处的温度可以通过例如以下等式来确定:

[0148] $T(t)=\alpha T(t-1),\alpha\in[0.85,0.96]$ (4)

[0149] 可以基于冷却时间表对等式(1)中所示的优先级函数的影响来选择冷却时间表。等式(1)中所示的优先级函数可以在温度高时优先添加新顶点,以增加弱连接组件中实体的数量。随着系统冷却,优先级函数可以在图形密度上施加更多权重,以便在图中已经存在的顶点之间添加更多边。优先级函数的两个部分可以通过一些系数(例如,Z和η)进行标准化,使得它们具有相似的数量级。

[0150] 图7图示了根据如上关于例如表2所述的某些实施例的基于用户数据集(例如,用户话语)生成的种子图700的示例。在该示例中,种子图700描述了Web应用及其父级公司之间的关系。种子图700可以基于RDF三元组,诸如<SoftwareA(软件A),part of(是……的一部分),CompanyA(公司A)>(从WikiData ID转换成WikiData标签)。如图所示,种子图700可以是具有很少实体710和在实体710之间很少的链接720的稀疏图。链接720被标记有实体710之间的关系。图7仅出于说明的目的而示出种子图700的稀疏性,并且种子图700的具体内容不旨在是易读的。

[0151] 图8图示了基于种子图700和以上关于例如表3描述的技术生成的定制知识图800的示例。图9是知识图800中的部分810的放大图。图8和图9仅出于说明的目的,并且知识图800的具体内容并非旨在是易读的。与种子图700相比,知识图800中可能有多得多的实体820和实体820之间的链接830。知识图800包括关于Web应用和周围概念的各种信息。一些示例包括:<Internet Message Access Protocol,part of,email>,<email filtering,instance of,anti-spam>,<RunKeeper,instance of,mobile app>,<RunKeeper,operating system,Android>和<CompanyA,country,United States of America>。知识图800还可以包括一些可能与数据集的领域不立即相关的信息,诸如,例如,<alternative rock,subclass of,rock music>和<calendar,instance of,chronology>。当提供不可见的数据时,潜在来自其它领域的此类信息可以使图更容易泛化。

[0152] 形成知识图800的RDF三元组的一些示例(格式为“主语|谓语|宾语”)可以包括“CompanyA Street View|developer|CompanyA”、“email filtering|instance of|anti-spam”、“BugMeNot|instance of|service on internet”、“CompanyB|country|United States of America”、“CompanyC.com|software engine|ProductC”、“RunKeeper|operating system|Android”、“RunKeeper|instance of|mobile app”、“ServiceA|instance of|online dating service”、“Internet Message Access Protocol|part of|email”、“software release life cycle|instance of|software engineering”、“CompanyD|owned by|CompanyE”、“CompanyF|parent organization|CompanyG”、“G Suite|industry|email”、“ProductA|publisher|CompanyH”、“ProductA|developer|CompanyH”等等。

[0153] 表4图示了使用参考知识图在扩展之前和之后的知识图的示例的统计结果。统计包括例如数据集中的顶点数、边数、图密度和弱连接组件的数量。如表4中所示,对于具有大量起始顶点的数据集,上述技术可能不添加许多新顶点,但可能会添加多得多的边。对于SNIPS,顶点数和边数都增加,但是图密度降低,因为SNIPS并不是有限的知识领域数据集。但是,对于较小的种子图,可能会添加大量的新顶点。这可能是因为,对于较大的数据集,为种子图识别的实体可能位于参考知识图上紧邻。此外,当起始顶点很多时,遍历参考知识图

时,可能不会发现很多新顶点,而是可能会发现这些实体之间的新关系。另一方面,对于较小的数据集以及因此较小的种子图,该技术更有可能发现更多新顶点。

[0154] 表4图统计

[0155]

数据集	|V|		|E|		图密度		计数(WCC)	
	之前	之后	之前	之后	之前	之后	之前	之后
SNIPS	13116	15223	30842	35160	1.794e-04	1.517e-04	3172	2955
Web 应用	173	206	69	117	2.319e-03	2.771e-03	127	114
聊天机器人	126	136	56	69	3.682e-03	3.758e-03	94	94
AskUbuntu	268	386	68	211	9.504e-04	1.420e-03	208	192

[0156] 知识图可以用于与自然语言理解相关的许多应用中。例如,知识图嵌入(KGE)技术可以将包括实体和关系的知识图的组件嵌入到连续的向量空间中,以便简化操纵,同时保留知识图的固有结构。然后,实体和关系嵌入可以用于例如知识图完成、关系提取、实体分类、实体解析、问题回答等。

[0157] 一般而言,KGE技术表示连续向量空间中的实体和关系,然后对每个事实和其它信息(例如,实体类型、关系路径、文本描述或逻辑规则)使用基于距离或相似性的评分函数来衡量其合理性。然后,可以通过使观察到的事实的总体合理性最大化来获得实体和关系嵌入。知识图嵌入可以比词向量包含更多的语义含义,并且可以用于基于词向量来增强特征表示。KGE可以通过例如捕获多词符号(例如,短语)的语义含义并提高带有词汇外(OOV)词(诸如,特别是在具有大量唯一且重要的词的小型数据集中没有词向量的词)的数据集的预测能力来捕获超出典型词向量的实体的语义表示。

[0158] 如上所述,聊天机器人是计算机程序或应用,其可以通过模拟人作为对话伙伴将如何表现来经由听觉或文本方法进行对话。聊天机器人可以在对话系统中用于各种实际目的,包括客户服务或信息获取。一些更简单的聊天机器人可以扫描输入中的关键字,然后从数据库中提取具有最佳匹配的关键字或最相似的措词模式的回复。一些聊天机器人可以使用更复杂的自然语言处理技术来更逼真地模拟人。

[0159] 根据某些实施例,知识图嵌入可以用于意图分类,其中任务可以包括例如由聊天机器人确定在给定输入自然语言话语下用户希望执行的动作。聊天机器人可以使用从例如使用上述方法生成的定制知识图提取的词向量和知识图嵌入两者,以改善用户意图分类的性能。

[0160] 根据某些实施例,诸如TransE技术之类的距离变换技术(参见,例如,Bordes等人,“Translating Embeddings for Modeling Multi-relational Data”,Advances in Neural Information Processing Systems 26.Curran Associates公司,2013年,第2787-2795页)及其扩展可以用于对WikiData上的嵌入进行预训练(参见例如Vrandecic& **Krötzsch** ,“Wikidata:A Free Collaborative Knowledgebase”,Communications of the ACM 57:10,2014年,第78-85页),因为基于变换的建模技术非常适合意图识别应用。TransE是用于学习实体的低维嵌入的基于能量的模型,其中关系表示为嵌入空间中的变换:如果(h,r,t)成立(例如,是正确的RDF),那么尾部实体t的嵌入靠近头部实体h的嵌入与依赖于关系r的向量的总和(t≈h+r);否则,尾部实体t的嵌入远离头部实体h的嵌入与依赖于关系r的向量的总和。

[0161] KGE技术(例如,基于距离变换或相似度的KGE技术)通常包括三个操作:(1)表示实体和关系,(2)定义评分函数,以及(3)学习实体和关系表示。第一个操作指定实体和关系在连续向量空间中表示的形式。实体通常表示为向量,诸如向量空间中的确定点。关系通常被视为向量空间中的操作,其可以表示为向量、矩阵、张量、多元高斯分布或高斯混合。在第二个操作中,将评分函数$f_r(h,t)$应用于每个事实(h,r,t),以衡量其合理性。例如,在TransE技术中,评分函数可以被定义为h+r与t之间的负距离:

[0162] $$f_r(h,t) = -\#h + r - t\#_{1/2}. \quad (5)$$

[0163] 知识图中观察到的事实通常比未观察到的事实分数更高。第三个操作涉及优化处理,该处理使观察到的事实的总合理性最大化,以学习实体和关系表示(例如,嵌入)。

[0164] 在一些实施例中,对于图G中的所有实体和对应关系,可以首先使用预训练的嵌入来初始化嵌入,然后使用诸如TransE技术和不同评分函数之类的技术来重新训练嵌入。例如,用于重新训练处理的评分函数可以从等式(5)更改为:

[0165] $$f_r(h,t) = -\theta_r\#h + r - t\#_{1/2}, \quad (6)$$

[0166] 其中θ_r是属于所生成的图的实体和关系的权重。这种加权评分函数是受TransM的启发(参见,例如,M.Fan等人,"Transition-based knowledge graph embedding with relational mapping proper-ties",in Proc.28th Pacific Asia Conf.Language Inf.Comput.,2014,第328-337页),并赋予图G中的实体和关系更大的权重。可以遵循TransE的开放世界训练范式使用最大余量铰链损失函数。可以根据以下梯度来更新嵌入:

[0167] $$\sum_{(\mathbf{h},\mathbf{r},\mathbf{t},)\in T}\sum_{(\mathbf{h'},\mathbf{r},\mathbf{t'})\in T'}[\gamma - f_r(\mathbf{h},\mathbf{t}) + f_{r'}(\mathbf{h'},\mathbf{t'}) + \lambda(\|\mathbf{r}\|_2^2)]_+ \quad \text{，或者} \quad (7)$$

$$\sum_{((h,r,t),(h',r',t'))\in T_{batch}} max(0,\gamma-(f_{(r)}(h+r,t)-f_{(r')}(h'+r',t'))+\lambda(h^2+t^2+r^2))$$

,

[0168] 其中$\#h\#_2=1$并且$\#t\#_2=1$。等式(7)可以包括最大余量铰链损失函数,其中超参数γ是余量(>0)。等式(7)中包含附加的正则项,以防止图G的实体和关系嵌入变得太稀疏,其中λ是正则强度(>0)。T和T'分别是指每个小批量中的训练三元组和损坏的三元组。这被建模为转移学习处理,其中图嵌入的重新训练可以通过将更大的权重赋予其中存在的实体和关系来微调所生成的知识图的向量。该处理可以允许通过优先考虑生成图中此类非一对一关系不太可能的构成来减少嵌入中的噪声并补偿非一对一的建模关系中TransE的弱点。

[0169] 表5示出了将如上所述的知识图嵌入用于例如意图分类的方法的示例。

[0170] 表5使用知识图嵌入进行意图分类

1. 使用 TransE 在 Wikidata 上训练（或使用预训练的）知识图嵌入
2. 对于从训练数据集（例如，用户数据集）生成的当前知识图中的所有实体和关系:
   a. 用预训练的嵌入初始化 TransE 算法
   b. 使用 TransE 重新训练嵌入，但用不同的评分函数
      1) 关于以下的梯度更新嵌入:

$$\sum_{((h,r,t),(h',r',t'))\in T_{batch}} max(0,\gamma-(f_{(r)}(h+r,t)-f_{(r')}(h'+r',t'))+\lambda(h^2+t^2+r^2))$$

,

[0171] 其中 h、r 和 t 分别是头部、关系和尾部实体向量，并且 h'、r'和 t'是损坏的头部、关系和尾部实体向量。其它变量是实常数。评分函数是:

$$f_r(h,t)=-\theta_r\|h+r-t\|_{1/2},$$

其中范式可以是例如欧几里得范式。新的评分函数在计算损失时会更重地对当前子图中的实体加权

3. 对查询嵌入求平均并与平均的 GloVe 向量组合
   a. 组合可以包括级联向量，然后对它们执行 PCA 以减小维度
4. 使用 GridSearch 将嵌入传递到逻辑回归模型中，以进行超参数调整和 5 倍交叉验证（例如，80:20 训练来测试拆分）

[0172] 在表5中所示的示例中,可以在操作1处在例如Wikidata上对嵌入进行预训练。在操作2处,对于知识图G中的所有实体和对应关系,可以用在操作1处获得的预训练的嵌入来初始化嵌入,然后可以使用诸如TransE技术和不同的加权的评分函数之类的技术对嵌入进行重新训练。在操作3处,对于给定的话语,可以使用上面关于表2描述的实体提取和链接处理来提取话语中的所有实体。然后,对句子中的实体的(如上计算出的)所有实体嵌入求平均。还可以并行清洗句子(例如,修剪、移除词汇、移除数字和未知字符等)以减少稀疏性,并且可以对句子中的剩余词的GloVe向量(参见例如,Pennington等人,“Glove:Global vectors for word representation”,Proceedings of the Empirical Methods in Natural Language Processing(EMNLP),2014)求平均。然后,将句子的平均GloVe向量和平均知识图嵌入进行级联,然后可以对级联的向量执行主组件分析(PCA),如在例如Goikoetxea等人的“Single or multiple?combining word representations independently learned from text and wordnet”,Proceedings of AAAI,2016中所示。在操作4处,PCA之后的级联向量可以传递给分类器以进行意图分类。

[0173] 表6图示了使用以上关于表5描述的知识图嵌入技术和平均GloVe向量技术的训练时间和查询时间。KGE训练时间用于没有高速缓存的首次运行。随后的运行将更快。

[0174] 表6训练和查询时间

[0175]

数据集	话语数量	KGE 训练时间	GloVe 训练时间	KGE 查询时间	GloVe 查询时间
SNIPS	13784	203 m	344 s	0.9 s	0.93 s
WebApplication	82	63 s	82 s	0.9 s	0.39 s
AskUbuntu	162	182 s	74 s	0.9 s	0.37 s

[0176] 还进行了消融研究,以确定整个处理的哪些方面可能有助于准确性和F1分数的提高。已经对该技术的五个变体进行了实验。在变体Var 1中,不执行超参数优化(即,不执行网格搜索),并且在一定范围内随机选择超参数。在变体Var 2中,知识图仅在输入话语中找到的实体之间通过顶点和仅边(关系)初始化。在变体Var 3中,使用最大深度为5的参考知识图的一跳随机游走遍历,而不是如本公开中所描述的有限状态机。在变体Var 4中,特征向量是通过对输入话语中每个词的GloVe向量求平均而生成的。在变体Var 5中,知识图嵌入未如上关于表5所述进行重新训练,而是使用为WikiData预训练的TransE嵌入。这五个变体的F1分数和准确性结果显示在表7中,并与未应用这些变体中的任何一个而获得的结果进行了比较。

[0177] 表7消融研究中的准确性数据

[0178]

数据集	完整算法		Var 1		Var 2		Var 3		Var 4		Var 5	
	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc
SNIPS	0.982	98.22%	0.979	97.86%	0.967	96.68%	0.969	96.88%	0.974	97.42%	0.970	97.02%
WebApplication	0.833	85.71%	0.818	84.61%	0.783	83.33%	0.741	81.94%	0.783	83.33%	0.818	84.61%
聊天机器人	0.976	97.62%	0.976	97.62%	0.966	96.60%	0.950	95.23%	0.94	93.98%	0.946	94.57%
AskUbuntu	0.901	90.12%	0.901	90.12%	0.880	88.00%	0.831	86.82%	0.876	87.59%	0.889	88.89%

[0179] 消融研究的结果表明,虽然超参数优化可能会在准确性上带来较小的改善(如表7中的Var 1与完整算法所示),但与本文公开的其它技术的贡献相比,它并不明显。随机游走遍历(变体Var 3)和平均GloVe向量(变体Var 4)的性能下降表明,知识图嵌入可以捕获更多与分类处理相关的信息。特别地,F1分数下降比准确性下降更明显。这可以指示在特征向量内可能存在更多关于话语的语义信息,其可以产生更少的假阳性和假阴性。初始化知识图的方法和遍历参考知识图的方法均会显著影响结果。重新训练知识图嵌入也可以提高性能,从而指示重新训练嵌入以有利于所生成的知识图的构成可以有助于对实体和关系进行建模。

[0180] 本文描述的技术可以在由一个或多个处理单元(例如,处理器、内核)、硬件或其组合执行的软件(例如,代码、指令、程序)中实现。软件可以存储在存储器中(例如,在诸如存储器设备之类的非暂态计算机可读存储介质上)。可以在包括云环境(例如,包括私有、公共和混合云环境的各种类型的云)、本地(on-premises)环境、混合环境等的各种不同环境中实现本文描述的技术。

[0181] 图10描绘了用于实现实施例的分布式系统1000的简化图。在所示的实施例中,分布式系统1000包括经由一个或多个通信网络1010耦合到服务器1012的一个或多个客户端计算设备1002、1004、1006和1008。客户端计算设备1002、1004、1006和1008可以被配置为执行一个或多个应用。

[0182] 在各种实施例中,服务器1012可以适于运行或提供一个或多个服务或软件应用,该服务或软件应用使得能够使用应用的需求信息在应用中的用户界面组件与数据对象之间自动生成关系或映射。例如,在某些实施例中,服务器1012可以接收应用的需求信息,其中需求信息可以包括对应用的期望功能的描述,并且还可以包括应用的期望GUI的图像。服务器1012然后可以使用需求信息来生成应用的用户界面组件与一个或多个数据对象之间的映射。

[0183] 在某些实施例中,服务器1012还可以提供可以在非虚拟和虚拟环境中提供的其它服务或软件应用。在一些实施例中,这些服务可以作为基于web的服务或云服务,诸如在软件即服务(SaaS)模型下,提供给客户端计算设备1002、1004、1006和/或1008的用户。操作客户端计算设备1002、1004、1006和/或1008的用户可以依次利用一个或多个客户端应用与服务器1012交互以利用由这些组件提供的服务。

[0184] 在图10所描绘的配置中,服务器1012可以包括实现由服务器1012执行的功能的一个或多个组件1018、1020和1022。这些组件可以包括可以由一个或多个处理器执行的软件组件、硬件组件或其组合。应当认识到的是,各种不同的系统配置是可能的,其可以与分布式系统1000不同。因此,图10中所示的实施例是用于实现实施例系统的分布式系统的一个示例,并且不旨在进行限制。

[0185] 根据本公开的教导,用户可以使用客户端计算设备1002、1004、1006和/或1008向服务器1012提供需求信息并从服务器1012接收信息(例如,映射信息)。客户端设备可以提供使客户端设备的用户能够与客户端设备交互的接口。客户端设备还可以经由该接口向用户输出信息。虽然图10仅描绘了四个客户端计算设备,但可以支持任何数量的客户端计算设备。

[0186] 客户端设备可以包括各种类型的计算系统,诸如便携式手持式设备、通用计算机,诸如个人计算机和膝上型计算机、工作站计算机、可穿戴设备、游戏系统、瘦客户端、各种消息传送设备、传感器和其它感测设备等。这些计算设备可以运行各种类型和版本的软件应用和操作系统(例如,Microsoft **Windows®**、Apple **Macintosh®**、**UNIX®**或类UNIX操作系统、Linux或类Linux操作系统,诸如Google Chrome™OS),包括各种移动操作系统(例如,Microsoft Windows **Mobile®** 、 **iOS®** 、Windows **Phone®** 、Android™、**BlackBerry®**,Palm **OS®**)。便携式手持式设备可以包括蜂窝电话、智能电话(例如,**iPhone®**)、平板电脑(例如,**iPad®**)、个人数字助理(PDA)等。可穿戴设备可以包括Google **Glass®**头戴式显示器和其它设备。游戏系统可以包括各种手持式游戏设备、具有互联网功能的游戏设备(例如,具有或不具有**Kinect®**手势输入设备的Microsoft **Xbox®**游戏机、Sony **PlayStation®**系统、由**Nintendo®**提供的各种游戏系统以及其它)等。客户端设备可以能够执行各种不同的应用,诸如各种与互联网相关的应用、通信应用(例如,电子邮件应用、短消息服务(SMS)应用),并且可以使用各种通信协议。

[0187] (一个或多个)通信网络1010可以是本领域技术人员熟悉的任何类型的网络,其可以使用多种可用协议中的任何一种来支持数据通信,包括但不限于TCP/IP(传输控制协议/

互联网协议)、SNA(系统网络体系架构)、IPX(互联网分组交换)、AppleTalk®等。仅仅作为示例,(一个或多个)通信网络1010可以是局域网(LAN)、基于以太网的网络、令牌环、广域网(WAN)、互联网、虚拟网络、虚拟专用网(VPN)、内联网、外联网、公共电话交换网(PSTN)、红外网络、无线网络(例如,在任何电气和电子协会(IEEE)1002.11协议套件、蓝牙®和/或任何其它无线协议下操作的网络)和/或这些和/或其它网络的任意组合。

[0188] 服务器1012可以由一个或多个通用计算机、专用服务器计算机(作为示例,包括PC(个人计算机)服务器、UNIX®服务器、中档服务器、大型计算机、机架安装的服务器等)、服务器场、服务器集群或任何其它适当的布置和/或组合组成。服务器1012可以包括运行虚拟操作系统的一个或多个虚拟机,或者涉及虚拟化的其它计算体系架构,诸如可以被虚拟化以维护服务器的虚拟存储设备的逻辑存储设备的一个或多个灵活的池。在各种实施例中,服务器1012可以适于运行提供前述公开中描述的功能的一个或多个服务或软件应用。

[0189] 服务器1012中的计算系统可以运行一个或多个操作系统,包括以上讨论的任何操作系统以及任何商用的服务器操作系统。服务器1012还可以运行各种附加服务器应用和/或中间层应用中的任何一种,包括HTTP(超文本传输协议)服务器、FTP(文件传输协议)服务器、CGI(通用网关接口)服务器、JAVA®服务器、数据库服务器等。示例性数据库服务器包括但不限于可从Oracle®、Microsoft®、Sybase®、IBM®(国际商业机器)等商购获得的数据库服务器。

[0190] 在一些实现中,服务器1012可以包括一个或多个应用以分析和整合从客户端计算设备1002、1004、1006和1008的用户接收到的数据馈送和/或事件更新。作为示例,数据馈送和/或事件更新可以包括但不限于从一个或多个第三方信息源和连续数据流接收到的Twitter®馈送、Facebook®更新或实时更新,其可以包括与传感器数据应用、金融报价机、网络性能测量工具(例如,网络监视和流量管理应用)、点击流分析工具、汽车流量监视等相关的实时事件。服务器1012还可以包括经由客户端计算设备1002、1004、1006和1008的一个或多个显示设备显示数据馈送和/或实时事件的一个或多个应用。

[0191] 分布式系统1000还可以包括一个或多个数据储存库1014、1016。在某些实施例中,这些数据储存库可以用于存储数据和其它信息。例如,一个或多个数据储存库1014、1016可以用于存储信息,诸如一个或多个应用的需求信息、映射信息,以及由服务器1012使用或生成的其它信息。数据储存库1014、1016可以驻留在各种位置。例如,由服务器1012使用的数据储存库可以在服务器1012本地,或者可以远离服务器1012并经由基于网络或专用的连接与服务器1012通信。数据储存库1014、1016可以是不同的类型。在某些实施例中,由服务器1012使用的数据储存库可以是数据库,例如关系数据库,诸如由Oracle®公司和其它供应商提供的数据库。这些数据库中的一个或多个可以适于响应于SQL格式的命令来实现数据库中数据的存储、更新和检索。由应用使用的数据储存库可以具有不同的类型,诸如,例如,键-值储存库、对象储存库或由文件系统支持的通用储存库。

[0192] 在某些实施例中,可以经由云环境将本公开中描述的特征作为服务来提供。图11是根据某些实施例的其中各种服务可以被提供为云服务的基于云的系统环境的简化框图。在图11所示的实施例中,云基础设施系统1102可以提供可以由用户使用一个或多个客户端

设备1104、1106和1108请求的一个或多个云服务。云基础设施系统1102可以包括一个或多个计算机和/或服务器,其可以包括以上针对服务器812描述的那些计算机和/或服务器。云基础设施系统1102中的计算机可以被组织为通用计算机、专用服务器计算机、服务器场、服务器集群或任何其它适当的布置和/或组合。

[0193] (一个或多个)网络1110可以促进客户端设备1104、1106和1108与云基础设施系统1102之间的通信和数据交换。(一个或多个)网络1110可以包括一个或多个网络。网络可以是相同或不同的类型。(一个或多个)网络1110可以支持一种或多种通信协议,包括有线和/或无线协议,以促进通信。

[0194] 图11中描绘的实施例仅仅是云基础设施系统的一个示例,并且不旨在进行限制。应该认识到的是,在一些其它实施例中,云基础设施系统1102可以具有比图11所示的组件更多或更少的组件、可以组合两个或更多个组件,或者可以具有不同的组件配置或布置。例如,虽然图11描绘了三个客户端计算设备,但是在替代实施例中可以支持任何数量的客户端计算设备。

[0195] 术语“云服务”通常用于指由服务提供商的系统(例如,云基础设施系统1102)根据需要并且经由诸如互联网之类的通信网络使用户可使用的服务。典型地,在公共云环境中,组成云服务提供商系统的服务器和系统与客户自己的内部服务器和系统不同。云服务提供商的系统由云服务提供商管理。客户因此可以利用由云服务提供商提供的云服务,而不必为服务购买单独的许可证、支持或硬件和软件资源。例如,云服务提供商的系统可以托管应用,并且用户可以经由互联网按需订购和使用应用,而用户不必购买用于执行应用的基础设施资源。云服务旨在提供对应用、资源和服务的轻松、可扩展的访问。几个提供商提供云服务。例如,由加利福尼亚州Redwood Shores的Oracle®公司提供了几种云服务,诸如中间件服务、数据库服务、Java云服务等。

[0196] 在某些实施例中,云基础设施系统1102可以使用诸如软件即服务(SaaS)模型、平台即服务(PaaS)模型、基础设施即服务(IaaS)模型以及包括混合服务模型的其它模型之类的不同模型提供一个或多个云服务。云基础设施系统1102可以包括一套应用、中间件、数据库以及使得能够供给各种云服务的其它资源。

[0197] SaaS模型使得应用或软件能够作为服务通过如互联网的通信网络交付给客户,而客户不必为底层应用购买硬件或软件。例如,SaaS模型可以用于为客户提供对由云基础设施系统1102托管的按需应用的访问。由Oracle®公司提供的SaaS服务的示例包括但不限于用于人力资源/资本管理、客户关系管理(CRM)、企业资源计划(ERP)、供应链管理(SCM)、企业绩效管理(EPM)、分析服务、社交应用及其它的各种服务。在某些实施例中,从应用的需求信息自动生成工作应用的能力可以作为SaaS模型下的服务提供。

[0198] IaaS模型通常用于向客户提供基础设施资源(例如,服务器、存储装置、硬件和联网资源)作为云服务,以提供弹性计算和存储能力。Oracle®公司提供了各种IaaS服务。

[0199] PaaS模型通常用于提供平台和环境资源作为服务,其使得客户能够开发、运行和管理应用和服务,而客户不必采购、构建或维护此类资源。由Oracle®公司提供的PaaS服务的示例包括但不限于Oracle Java云服务(JCS)、Oracle数据库云服务(DBCS)、数据管理云服务、各种应用开发解决方案服务以及其它服务。

[0200] 云服务通常基于按需自助服务、基于订阅、弹性可缩放、可靠、高度可用和安全的

方式提供。例如,客户可以经由订阅订单订购由云基础设施系统1102提供的一个或多个服务。然后,云基础设施系统1102执行处理,以提供客户的订阅订单中所请求的服务。例如,用户可以向云基础设施系统提供应用的需求信息,并请求云基础设施系统根据需求信息自动生成工作应用,包括如上所述的自动生成应用的用户界面组件和数据对象之间的映射。云基础设施系统1102可以被配置为提供一个或多个云服务。

[0201] 云基础设施系统1102可以经由不同的部署模型来提供云服务。在公共云模型中,云基础设施系统1102可以由第三方云服务提供商拥有,并且云服务被提供给任何普通公众客户,其中客户可以是个人或企业。在某些其它实施例中,在私有云模型下,可以在组织内(例如,在企业组织内)操作云基础设施系统1102,并向组织内的客户提供服务。例如,客户可以是企业的各个部门,诸如人力资源部门、工资部门等,甚至是企业内的个人。在某些其它实施例中,在社区云模型下,云基础设施系统1102和所提供的服务可以由相关社区中的几个组织共享。也可以使用各种其它模型,诸如上面提到的模型的混合。

[0202] 客户端设备1104、1106和1108可以是不同类型的(诸如图11中描绘的客户端设备1104、1106和1108),并且可以能够操作一个或多个客户端应用。用户可以使用客户端设备与云基础设施系统1102交互,诸如请求由云基础设施系统1102提供的服务。例如,用户可以使用客户端设备向云基础设施系统1102提供应用的需求信息,并请求云基础设施系统根据需求信息自动生成工作应用,包括如本公开中所述的自动生成应用的用户界面组件和数据对象之间的映射。

[0203] 在一些实施例中,由云基础设施系统1102执行的用于提供所请求的服务的处理可能涉及大数据分析。这种分析可能涉及使用、分析和操纵大型数据集,以检测和可视化数据内的各种趋势、行为、关系等。该分析可以由一个或多个处理器执行、可能并行处理数据、使用数据执行仿真等。用于该分析的数据可以包括结构化数据(例如,存储在数据库中或根据结构化模型结构化的数据)和/或非结构化数据(例如,数据Blob(二进制大对象))。

[0204] 如在图11的实施例中所描绘的,云基础设施系统1102可以包括基础设施资源1130,其用于促进由云基础设施系统1102提供的各种云服务的供给。基础设施资源1130可以包括例如处理资源、存储或存储器资源、联网资源等。

[0205] 在某些实施例中,为了促进这些资源的高效供给以支持由云基础设施系统1102为不同客户提供的各种云服务,可以将资源捆绑成资源或资源模块的集合(也称为“群聚(pod)”)。每个资源模块或群聚可以包括一种或多种类型的资源的预先集成和优化的组合。在某些实施例中,可以为不同类型的云服务预先供给不同的群聚。例如,可以为数据库服务供给第一组群聚,可以为Java服务供给第二组群聚,其中第二组群聚可以包括与第一组群聚中的群聚不同的资源组合。对于一些服务,可以在服务之间共享为供给服务而分配的资源。

[0206] 云基础设施系统1102本身可以内部使用服务1132,服务1132由云基础设施系统1102的不同组件共享并且促进云基础设施系统1102的服务供给。这些内部共享的服务可以包括但不限于安全和身份服务、集成服务、企业储存库服务、企业管理器服务、病毒扫描和白名单服务、高可用性、备份和恢复服务、用于启用云支持的服务、电子邮件服务、通知服务、文件传输服务等。

[0207] 云基础设施系统1102可以包括多个子系统。这些子系统可以用软件或硬件或其组

合来实现。如图11所示,子系统可以包括用户界面子系统1112,该用户界面子系统1112使得云基础设施系统1102的用户或客户能够与云基础设施系统1102交互。用户界面子系统1112可以包括各种不同的界面,诸如web界面1114、在线商店界面1116(其中由云基础设施系统1102提供的云服务被广告并且可由消费者购买)和其它界面1118。例如,客户可以使用客户端设备通过使用接口1114、1116和1118中的一个或多个下订阅订单来请求(服务请求1134)由云基础设施系统1102提供的一个或多个服务。这些界面可以包括例如在线商店用户界面1116,客户访问和浏览在线商店,识别由云基础设施系统1102提供的云服务,并为提供的服务中客户希望订阅的一个或多个服务下订阅订单。其它界面可以包括web界面1114和其它界面1118。

[0208] 在某些实施例中,服务请求或订阅请求可以包括识别请求用户或客户以及客户期望订阅的一个或多个服务的信息。例如,客户可以为由云基础设施系统1102提供的自动化应用生成服务下订阅订单。

[0209] 在某些实施例中,诸如图11所示的实施例,云基础设施系统1102可以包括被配置为处理新的订阅订单的订单管理子系统(OMS)1120。作为该处理的一部分,OMS1120可以被配置为:为客户创建账户(如果尚未创建);接收来自客户的账单和/或会计信息,该账单和/或账单信息将用于针对向客户提供所请求的服务向客户计费;验证客户信息;在验证后,为客户预订订单;已经编排各种工作流程以准备用于供给的订单。

[0210] 一旦被正确地证实,OMS1120就可以调用订单供给子系统(OPS)1124,其被配置为为订单供给资源,包括处理资源、存储器资源和联网资源。供给可以包括为订单分配资源,以及配置资源以促进由客户订单所请求的服务。为订单供给资源的方式和供给资源的类型可以取决于客户已订购的云服务的类型。例如,根据一个工作流程,OPS1124可以被配置为确定所请求的特定云服务,并且识别可能已经针对该特定云服务而被预先配置的多个群聚。为订单分配的群聚的数量可以取决于所请求的服务的大小/数量/级别/范围。例如,可以基于服务所支持的用户的数量、正在请求的服务的持续时间等来确定要分配的群聚的数量。然后,可以针对特定的请求客户定制所分配的群聚,以提供所请求的服务。

[0211] 云基础设施系统1102可以向请求客户发送响应或通知1144,以指示所请求的服务何时准备就绪。在一些情况下,可以将信息(例如,链接)发送给客户,使得客户能够开始使用和利用所请求的服务的益处。

[0212] 云基础设施系统1102可以向多个客户提供服务。对于每个客户,云基础设施系统1102负责管理与从客户接收到的一个或多个订阅订单相关的信息、维护与订单相关的客户数据,以及向客户提供所请求的服务。云基础设施系统1102还可以收集关于客户对已订阅的服务的使用的使用统计信息。例如,可以针对使用的存储量、传输的数据量、用户的数量以及系统正常运行时间和系统停机时间量等收集统计信息。该使用信息可以用于向客户计费。计费可以例如按月周期进行。

[0213] 云基础设施系统1102可以并行地向多个客户提供服务。云基础设施系统1102可以存储这些客户的信息,包括可能的专有信息。在某些实施例中,云基础设施系统1102包括身份管理子系统(IMS)1128,其被配置为管理客户的信息并提供所管理的信息的分离,使得与一个客户相关的信息不可被另一个客户访问或与另一个客户的信息混合。IMS1128可以被配置为提供各种与安全相关的服务,诸如身份服务,诸如信息访问管理、认证和授权服务、

用于管理客户身份和角色及相关能力的服务等。

[0214] 图12图示了可以用于实现某些实施例的示例性计算机系统1200。例如,在一些实施例中,计算机系统1200可以用于实现上述各种服务器和计算机系统中的任何一个,诸如图1中描绘的模型生成系统122。如图12所示,计算机系统1200包括各种子系统,包括经由总线子系统1202与多个其它子系统通信的处理子系统1204。这些其它子系统可以包括处理加速单元1206、I/O子系统1208、存储子系统1218和通信子系统1224。存储子系统1218可以包括非暂态计算机可读存储介质,其包括存储介质1222和系统存储器1210。

[0215] 总线子系统1202提供用于使计算机系统1200的各种组件和子系统按照期望彼此通信的机制。虽然总线子系统1202被示意性地示为单条总线,但是总线子系统的替代实施例可以利用多条总线。总线子系统1202可以是几种类型的总线结构中的任何一种,包括存储器总线或存储器控制器、外围总线、使用任何各种总线体系架构的局部总线等。例如,此类体系架构可以包括工业标准体系架构(ISA)总线、微通道体系架构(MCA)总线、增强型ISA(EISA)总线、视频电子标准协会(VESA)局部总线和外围组件互连(PCI)总线,其可以实现为根据IEEE P1386.1标准制造的夹层(Mezzanine)总线,等等。

[0216] 处理子系统1204控制计算机系统1200的操作,并且可以包括一个或多个处理器、专用集成电路(ASIC)或现场可编程门阵列(FPGA)。处理器可以包括单核或多核处理器。可以将计算机系统1200的处理资源组织成一个或多个处理单元1232、1234等。处理单元可以包括一个或多个处理器、来自相同或不同处理器的一个或多个核心、核心和处理器的组合、或核心和处理器的其它组合。在一些实施例中,处理子系统1204可以包括一个或多个专用协处理器,诸如图形处理器、数字信号处理器(DSP)等。在一些实施例中,处理子系统1204的一些或全部可以使用定制电路来实现,诸如专用集成电路(ASIC)或现场可编程门阵列(FPGA)。

[0217] 在一些实施例中,处理子系统1204中的处理单元可以执行存储在系统存储器1210中或计算机可读存储介质1222上的指令。在各个实施例中,处理单元可以执行各种程序或代码指令,并且可以维护多个并发执行的程序或进程。在任何给定的时间,要执行的程序代码中的一些或全部可以驻留在系统存储器1210中和/或计算机可读存储介质1222上,包括可能在一个或多个存储设备上。通过适当的编程,处理子系统1204可以提供上述各种功能。在计算机系统1200正在执行一个或多个虚拟机的情况下,可以将一个或多个处理单元分配给每个虚拟机。

[0218] 在某些实施例中,可以可选地提供处理加速单元1206,以用于执行定制的处理或用于卸载由处理子系统1204执行的一些处理,从而加速由计算机系统1200执行的整体处理。

[0219] I/O子系统1208可以包括用于向计算机系统1200输入信息和/或用于从或经由计算机系统1200输出信息的设备和机制。一般而言,术语“输入设备”的使用旨在包括用于向计算机系统1200输入信息的所有可能类型的设备和机制。用户界面输入设备可以包括,例如,键盘、诸如鼠标或轨迹球之类的指向设备、并入到显示器中的触摸板或触摸屏、滚轮、点击轮、拨盘、按钮、开关、小键盘、带有语音命令识别系统的音频输入设备、麦克风以及其它类型的输入设备。用户界面输入设备还可以包括使用户能够控制输入设备并与之交互的诸如Microsoft **Kinect®**运动传感器的运动感测和/或姿势识别设备、Microsoft **Xbox®**

360游戏控制器、提供用于接收使用姿势和口语命令的输入的界面的设备。用户界面输入设备还可以包括眼睛姿势识别设备，诸如从用户检测眼睛活动(例如，当拍摄图片和/或进行菜单选择时的“眨眼”)并将眼睛姿势转换为到输入设备(例如，Google**Glass®**)的输入的Google**Glass®**眨眼检测器。此外，用户界面输入设备可以包括使用户能够通过语音命令与语音识别系统(例如，**Siri®**导航器)交互的语音识别感测设备。

[0220] 用户界面输入设备的其它示例包括但不限于，三维(3D)鼠标、操纵杆或指示杆、游戏板和图形平板、以及音频/视频设备，诸如扬声器、数字相机、数字摄像机、便携式媒体播放器、网络摄像机、图像扫描仪、指纹扫描仪、条形码读取器3D扫描仪、3D打印机、激光测距仪、以及眼睛注视跟踪设备。此外，用户界面输入设备可以包括，例如，医疗成像输入设备，诸如计算机断层摄影、磁共振成像、位置发射断层摄影、以及医疗超声检查设备。用户界面输入设备也可以包括，例如，音频输入设备，诸如MIDI键盘、数字乐器等。

[0221] 一般而言，术语“输出设备”的使用旨在包括所有可能类型的设备和用于从计算机系统1200向用户或其它计算机输出信息的机制。用户界面输出设备可以包括显示子系统、指示器灯或诸如音频输出设备的非可视显示器等。显示子系统可以是阴极射线管(CRT)、诸如利用液晶显示器(LCD)或等离子体显示器的平板设备、投影设备、触摸屏等。例如，用户界面输出设备可以包括但不限于，可视地传达文本、图形和音频/视频信息的各种显示设备，诸如监视器、打印机、扬声器、耳机、汽车导航系统、绘图仪、语音输出设备和调制解调器。

[0222] 存储子系统1218提供用于存储由计算机系统1200使用的信息和数据的储存库或数据存储库。存储子系统1218提供了用于存储提供某些实施例的功能的基本编程和数据构造的有形非暂态计算机可读存储介质的示例。存储子系统1218可以存储软件(例如，程序、代码模块、指令)，该软件在由处理子系统1204执行时提供上述功能。软件可以由处理子系统1204的一个或多个处理单元执行。存储子系统1218还可以提供用于存储根据本公开的教导使用的数据的储存库。

[0223] 存储子系统1218可以包括一个或多个非暂态存储器设备，包括易失性和非易失性存储器设备。如图12所示，存储子系统1218包括系统存储器1210和计算机可读存储介质1222。系统存储器1210可以包括多个存储器，包括用于在程序执行期间存储指令和数据的易失性主随机存取存储器(RAM)以及其中存储有固定指令的非易失性只读存储器(ROM)或闪存。在一些实现中，基本输入/输出系统(BIOS)可以典型地存储在ROM中，该基本输入/输出系统(BIOS)包含有助于例如在启动期间在计算机系统1200内的元件之间传递信息的基本例程。RAM通常包含当前由处理子系统1204操作和执行的数据和/或程序模块。在一些实现中，系统存储器1210可以包括多种不同类型的存储器，诸如静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)等。

[0224] 作为示例而非限制，如图12所示，系统存储器1210可以加载正在被执行的可以包括各种应用(诸如Web浏览器、中间层应用、关系型数据库管理系统(RDBMS)等)的应用程序1212、程序数据1214和操作系统1216。作为示例，操作系统1216可以包括各种版本的Microsoft **Windows®**、Apple**Macintosh®**和/或Linux操作系统、各种商用**UNIX®**或类UNIX操作系统(包括但不限于各种GNU/Linux操作系统、Google**Chrome®**OS等)和/或移动操作系统，诸如**iOS®**、**Windows®** Phone、**Android®** OS、**BlackBerry®** OS、

Palm®OS操作系统,以及其它操作系统。

[0225] 计算机可读存储介质1222可以存储提供一些实施例的功能的编程和数据构造。计算机可读存储介质1222可以为计算机系统1200提供计算机可读指令、数据结构、程序模块和其它数据的存储。当由处理子系统1204执行时,提供上述功能的软件(程序、代码模块、指令)可以存储在存储子系统1218中。作为示例,计算机可读存储介质1222可以包括非易失性存储器,诸如硬盘驱动器、磁盘驱动器、诸如CD ROM、DVD、**Blu-Ray®** (蓝光)盘或其它光学介质的光盘驱动器。计算机可读存储介质1222可以包括但不限于,**Zip®**驱动器、闪存存储器卡、通用串行总线(USB)闪存驱动器、安全数字(SD)卡、DVD盘、数字视频带等。计算机可读存储介质1222也可以包括基于非易失性存储器的固态驱动器(SSD)(诸如基于闪存存储器的SSD、企业闪存驱动器、固态ROM等)、基于易失性存储器的SSD(诸如基于固态RAM、动态RAM、静态RAM、DRAM的SSD、磁阻RAM(MRAM)SSD)以及使用基于DRAM和基于闪存存储器的SSD的组合的混合SSD。

[0226] 在某些实施例中,存储子系统1218还可以包括计算机可读存储介质读取器1220,其还可以连接到计算机可读存储介质1222。计算机可读存储介质读取器1220可以接收并且被配置为从诸如盘、闪存驱动器等的存储器设备读取数据。

[0227] 在某些实施例中,计算机系统1200可以支持虚拟化技术,包括但不限于处理和存储器资源的虚拟化。例如,计算机系统1200可以提供用于执行一个或多个虚拟机的支持。在某些实施例中,计算机系统1200可以执行诸如促进虚拟机的配置和管理的管理程序之类的程序。可以为每个虚拟机分配存储器、计算(例如,处理器、核心)、I/O和联网资源。每个虚拟机通常独立于其它虚拟机运行。虚拟机通常运行其自己的操作系统,该操作系统可以与由计算机系统1200执行的其它虚拟机执行的操作系统相同或不同。因此,计算机系统1200可以潜在地同时运行多个操作系统。

[0228] 通信子系统1224提供到其它计算机系统和网络的接口。通信子系统1224用作用于从计算机系统1200接收数据以及向其它系统传输数据的接口。例如,通信子系统1224可以使得计算机系统1200能够经由互联网建立到一个或多个客户端设备的通信信道,以用于从客户端设备接收信息以及向客户端设备发送信息。

[0229] 通信子系统1224可以支持有线和/或无线通信协议两者。例如,在某些实施例中,通信子系统1224可以包括用于(例如,使用蜂窝电话技术、高级数据网络技术(诸如3G、4G或EDGE(全球演进的增强数据速率)、WiFi(IEEE 802.XX族标准)、或其它移动通信技术、或其任意组合)接入无线语音和/或数据网络的射频(RF)收发器组件、全球定位系统(GPS)接收器组件和/或其它组件。在一些实施例中,作为无线接口的附加或替代,通信子系统1224可以提供有线网络连接(例如,以太网)。

[0230] 通信子系统1224可以以各种形式接收和传输数据。例如,在一些实施例中,除了其它形式之外,通信子系统1224还可以以结构化和/或非结构化的数据馈送1226、事件流1228、事件更新1230等形式接收输入通信。例如,通信子系统1224可以被配置为实时地从社交媒体网络的用户和/或诸如**Twitter®**馈送、**Facebook®**更新、诸如丰富站点摘要(RSS)馈送的web馈送的其它通信服务接收(或发送)数据馈送1226,和/或来自一个或多个第三方信息源的实时更新。

[0231] 在某些实施例中,通信子系统1224可以被配置为以连续数据流的形式接收本质上可能是连续的或无界的没有明确结束的数据,其中连续数据流可以包括实时事件的事件流1228和/或事件更新1230。生成连续数据的应用的示例可以包括例如传感器数据应用、金融报价机、网络性能测量工具(例如网络监视和流量管理应用)、点击流分析工具、汽车流量监视等。

[0232] 通信子系统1224也可以被配置为将数据从计算机系统1200传送到其它计算机系统或网络。数据可以以各种不同的形式传送给一个或多个数据库,诸如结构化和/或非结构化数据馈送1226、事件流1228、事件更新1230等,该一个或多个数据库可以与耦合到计算机系统1200的一个或多个流传输数据源进行通信。

[0233] 计算机系统1200可以是各种类型中的一种,包括手持便携式设备(例如,iPhone®蜂窝电话、iPad®计算平板、PDA)、可穿戴设备(例如,Google Glass®头戴式显示器)、个人计算机、工作站、大型机、信息站、服务器机架或任何其它数据处理系统。由于计算机和网络不断变化的性质,对图12中绘出的计算机系统1200的描述旨在仅仅作为具体示例。具有比图12中所绘出的系统更多或更少组件的许多其它配置是可能的。基于本文所提供的公开内容和教导,本领域普通技术人员将理解实现各种实施例的其它方式和/或方法。

[0234] 虽然已经描述了特定的实施例,但是各种修改、变更、替代构造以及等同物都是可能的。实施例不限于在某些特定数据处理环境内的操作,而是可以在多个数据处理环境内自由操作。此外,虽然已经使用一系列特定的事务和步骤描述了某些实施例,但是对于本领域技术人员来说清楚的是,这并不旨在进行限制。虽然一些流程图将操作描述为顺序处理,但是许多操作可以并行或同时执行。此外,操作的次序可以被重新布置。处理可能具有图中未包括的其它步骤。上述实施例的各种特征和方面可以被单独使用或联合使用。

[0235] 另外,虽然已经使用硬件和软件的特定组合描述了某些实施例,但是应该认识到的是,硬件和软件的其它组合也是可能的。某些实施例可以仅用硬件或仅用软件或其组合来实现。本文描述的各种处理可以以任何组合在相同的处理器或不同的处理器上实现。

[0236] 在将设备、系统、组件或模块描述为被配置为执行某些操作或功能的情况下,这样的配置可以通过以下方式来实现,例如,通过设计电子电路来执行操作、通过对可编程电子电路(诸如微处理器)进行编程来执行操作,诸如通过执行计算机指令或代码,或处理器或核心被编程为执行存储在非暂态存储介质上的代码或指令,或其任意组合来执行操作。进程可以使用各种技术进行通信,包括但不限于用于进程间通信的常规技术,并且不同对的进程可以使用不同的技术,或者同一对进程可以在不同时间使用不同的技术。

[0237] 在本公开中给出了具体细节以提供对实施例的透彻理解。但是,可以在没有这些具体细节的情况下实践实施例。例如,已经示出了众所周知的电路、处理、算法、结构和技术,而没有不必要的细节,以避免使实施例模糊。本描述仅提供示例实施例,并且不旨在限制其它实施例的范围、适用性或配置。相反,实施例的先前描述将为本领域技术人员提供用于实现各种实施例的使能描述。可以对元件的功能和布置进行各种改变。

[0238] 因此,说明书和附图应被认为是说明性的而不是限制性的。但是,将清楚的是,在不脱离权利要求书所阐述的更广泛的精神和范围的情况下,可以对其进行添加、减少、删除以及其它修改和改变。因此,虽然已经描述了具体的实施例,但是这些实施例并不旨在进行限制。各种修改和等同形式均在所附权利要求的范围内。

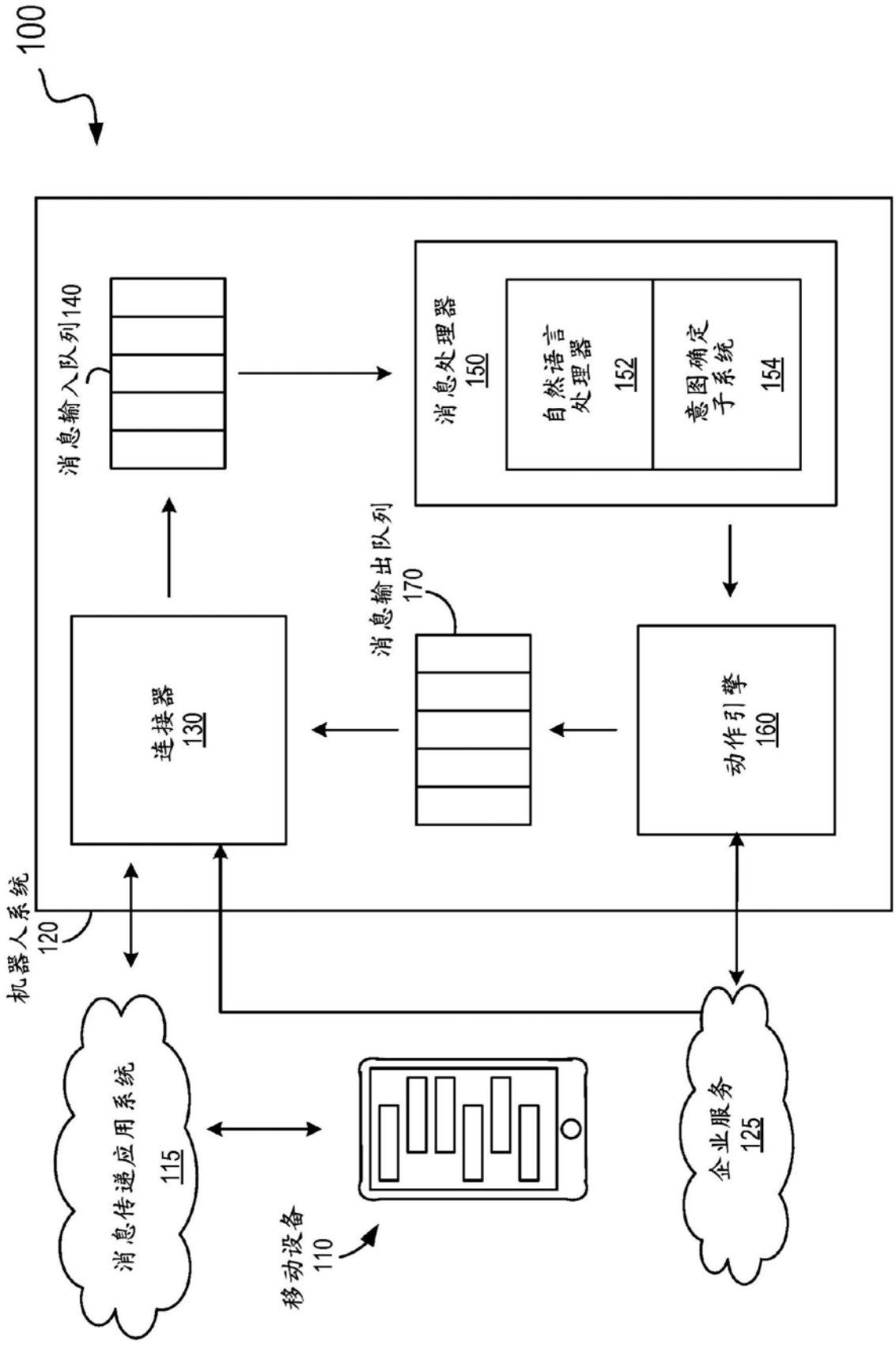

100
机器人系统
120
消息输入队列140
消息处理器
150
自然语言
处理器
152
意图确定
子系统
154
连接器
130
消息输出队列
170
动作引擎
160
消息传递应用系统
115
移动设备
110
企业服务
125

图1

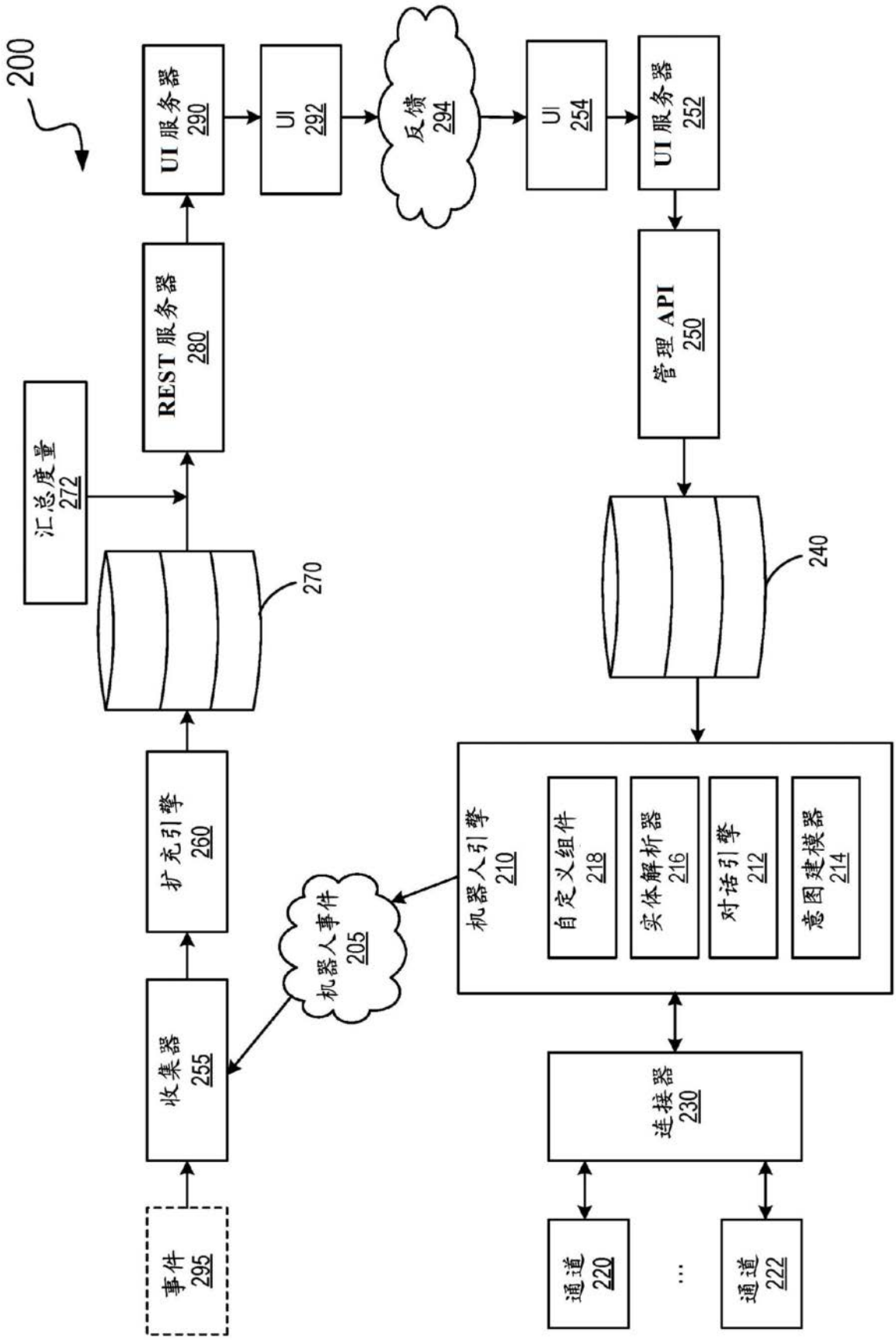

200
UI 服务器 290
UI 292
反馈 294
UI 254
UI 服务器 252
REST 服务器 280
管理 API 250
汇总度量 272
270
240
扩充引擎 260
机器人引擎 210
自定义组件 218
实体解析器 216
对话引擎 212
意图建模器 214
机器人事件 205
收集器 255
连接器 230
事件 295
通道 220
...
通道 222

图2

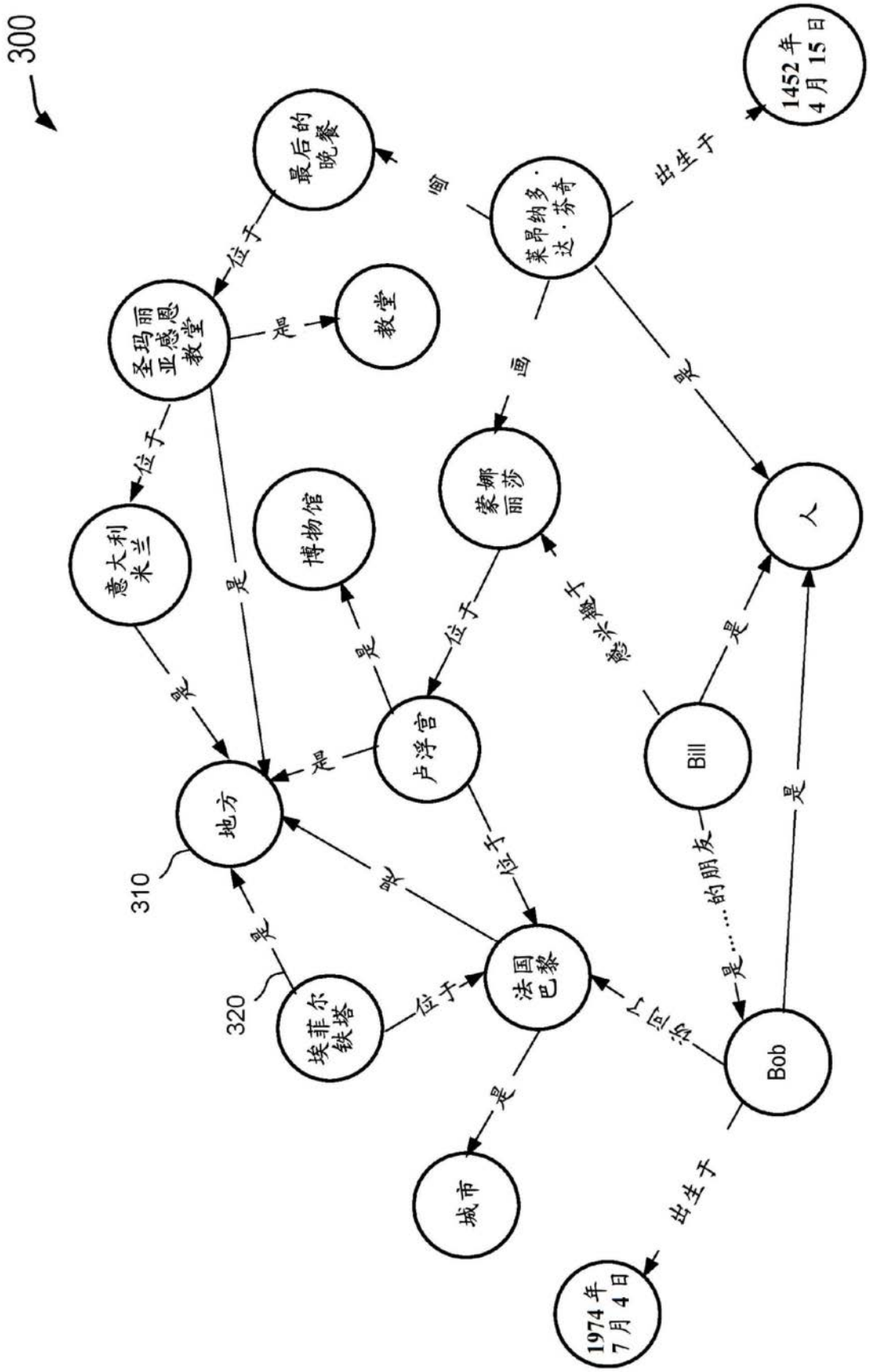

300
最后的晚餐
1452年4月15日
位于
画
出生于
圣玛丽亚感恩教堂
教堂
冲
莱昂纳多·达·芬奇
位于
画
意大利米兰
博物馆
蒙娜丽莎
人
是
是
位于
被
卢浮宫
冲
Bill
被
地方
310
是
320
吸
位于
法国巴黎
埃菲尔铁塔
位于
人
是……的朋友
Bob
是
是
城市
出生于
1974年7月4日

图3

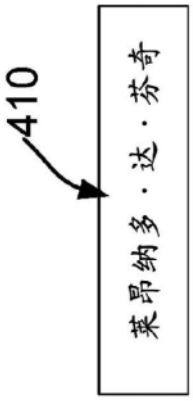


图4A

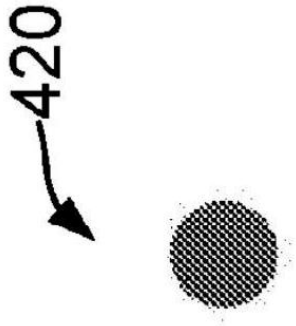


图4B

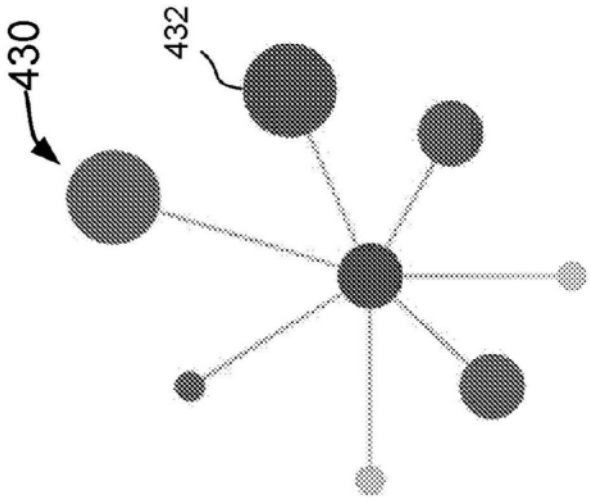


图4C

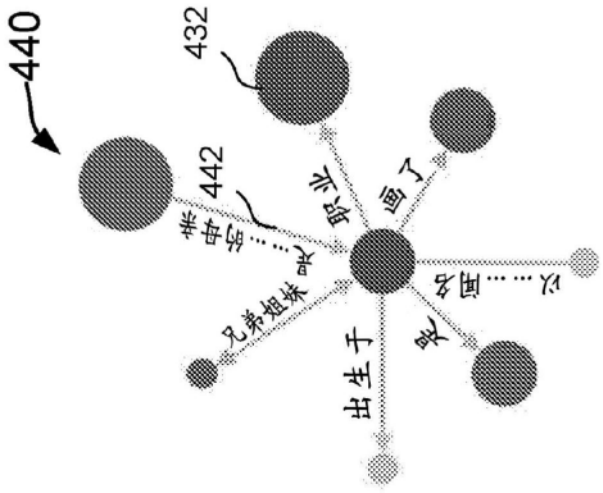


图4D

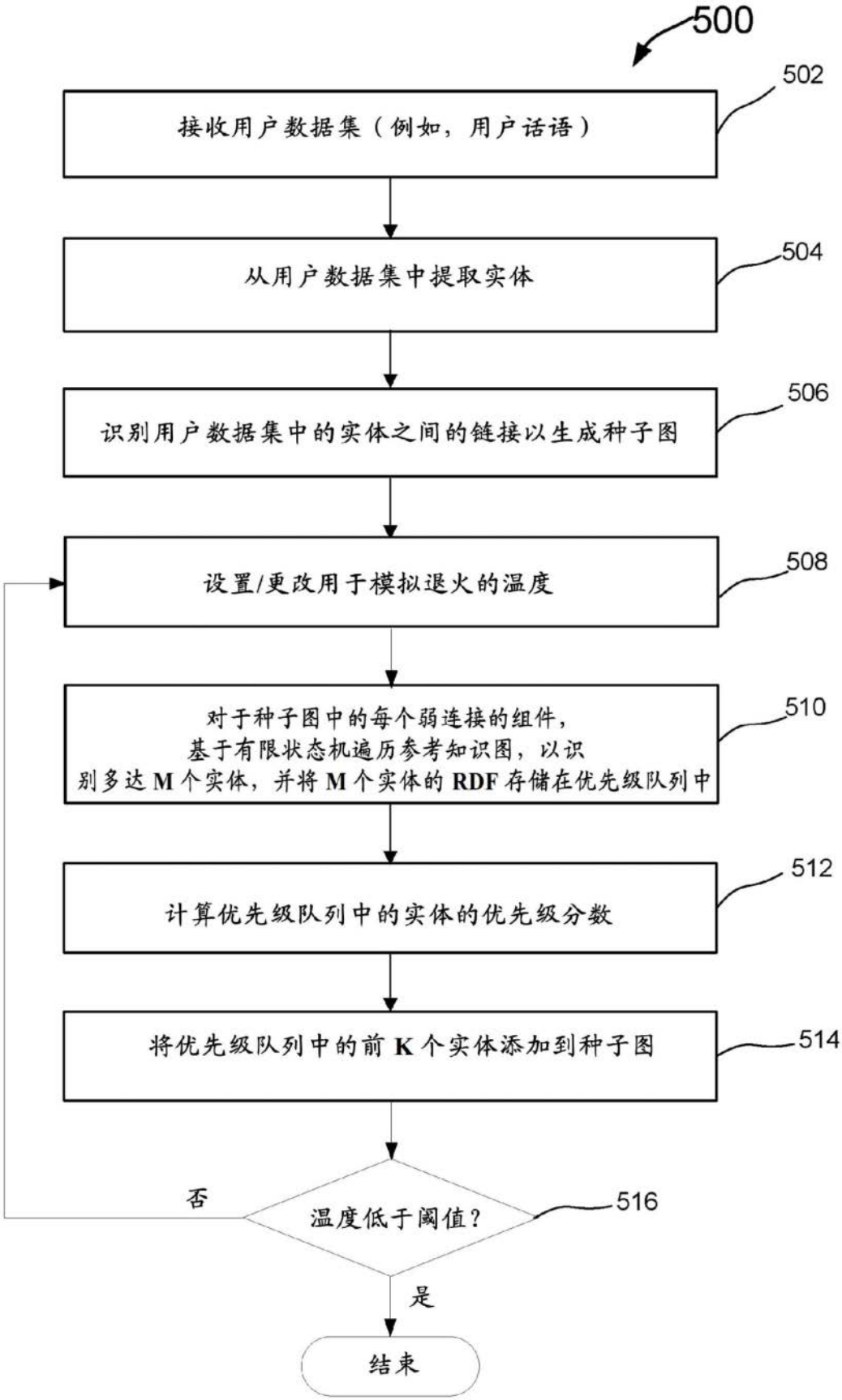

500
502
接收用户数据集（例如，用户话语）
504
从用户数据集中提取实体
506
识别用户数据集中的实体之间的链接以生成种子图
508
设置/更改用于模拟退火的温度
510
对于种子图中的每个弱连接的组件，
基于有限状态机遍历参考知识图，以识
别多达 M 个实体，并将 M 个实体的 RDF 存储在优先级队列中
512
计算优先级队列中的实体的优先级分数
514
将优先级队列中的前 K 个实体添加到种子图
516
温度低于阈值？
否
是
结束

图5

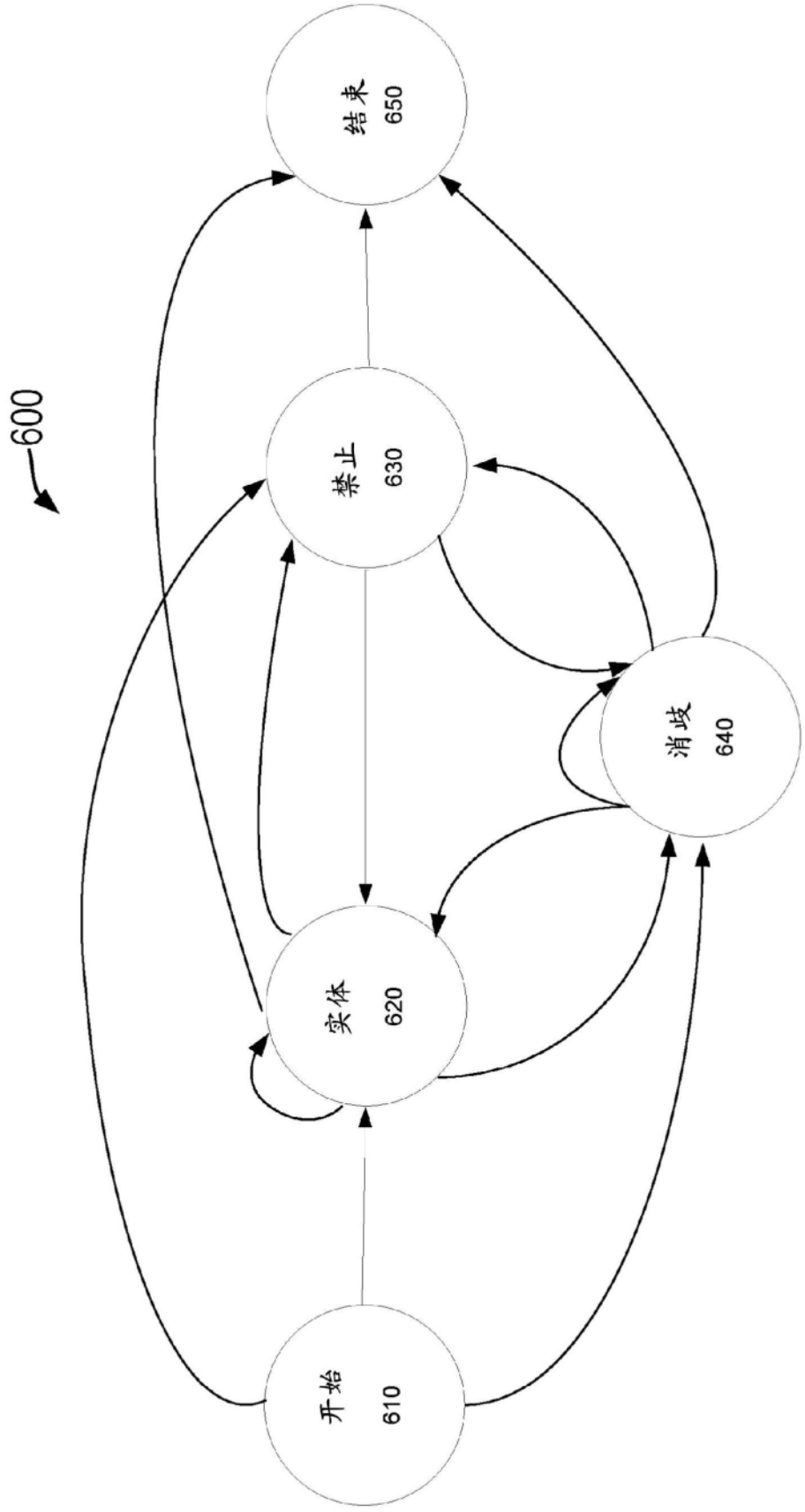


图6

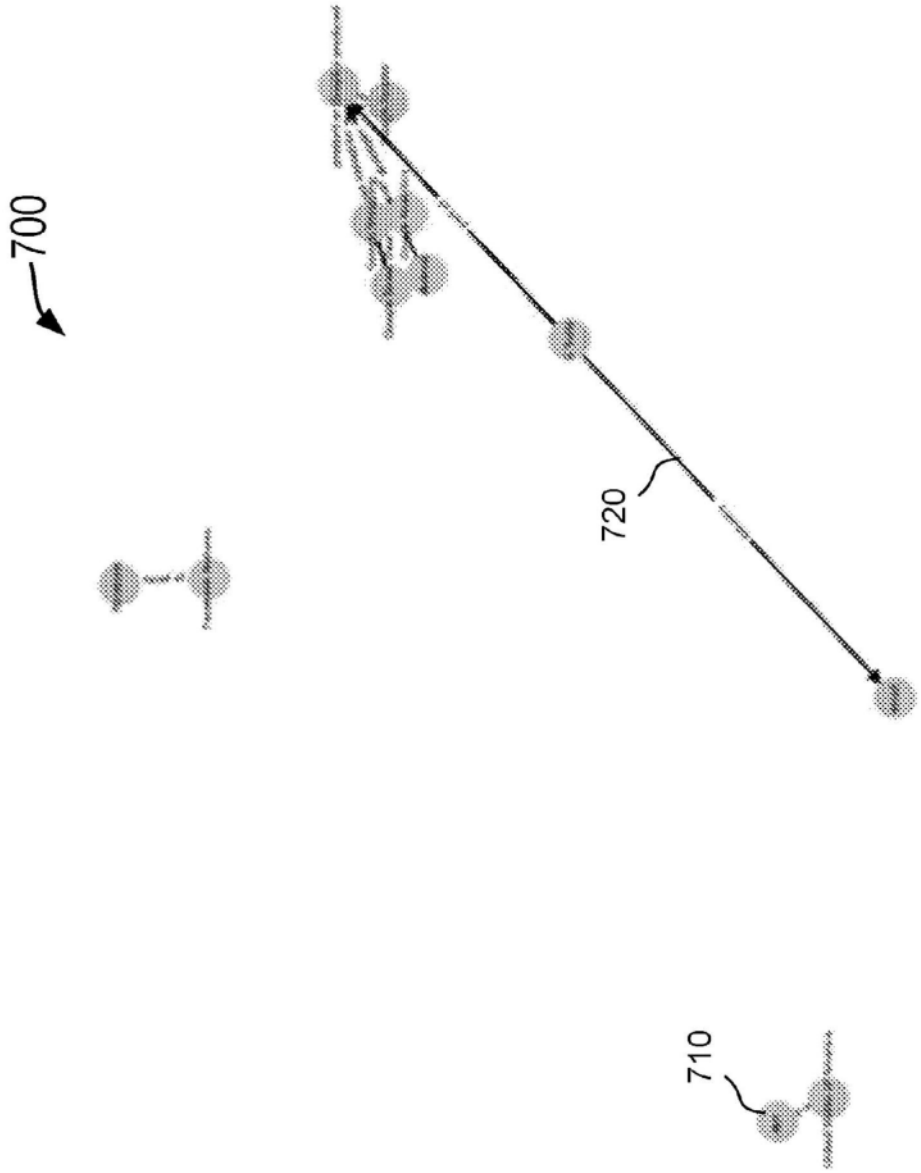


图7

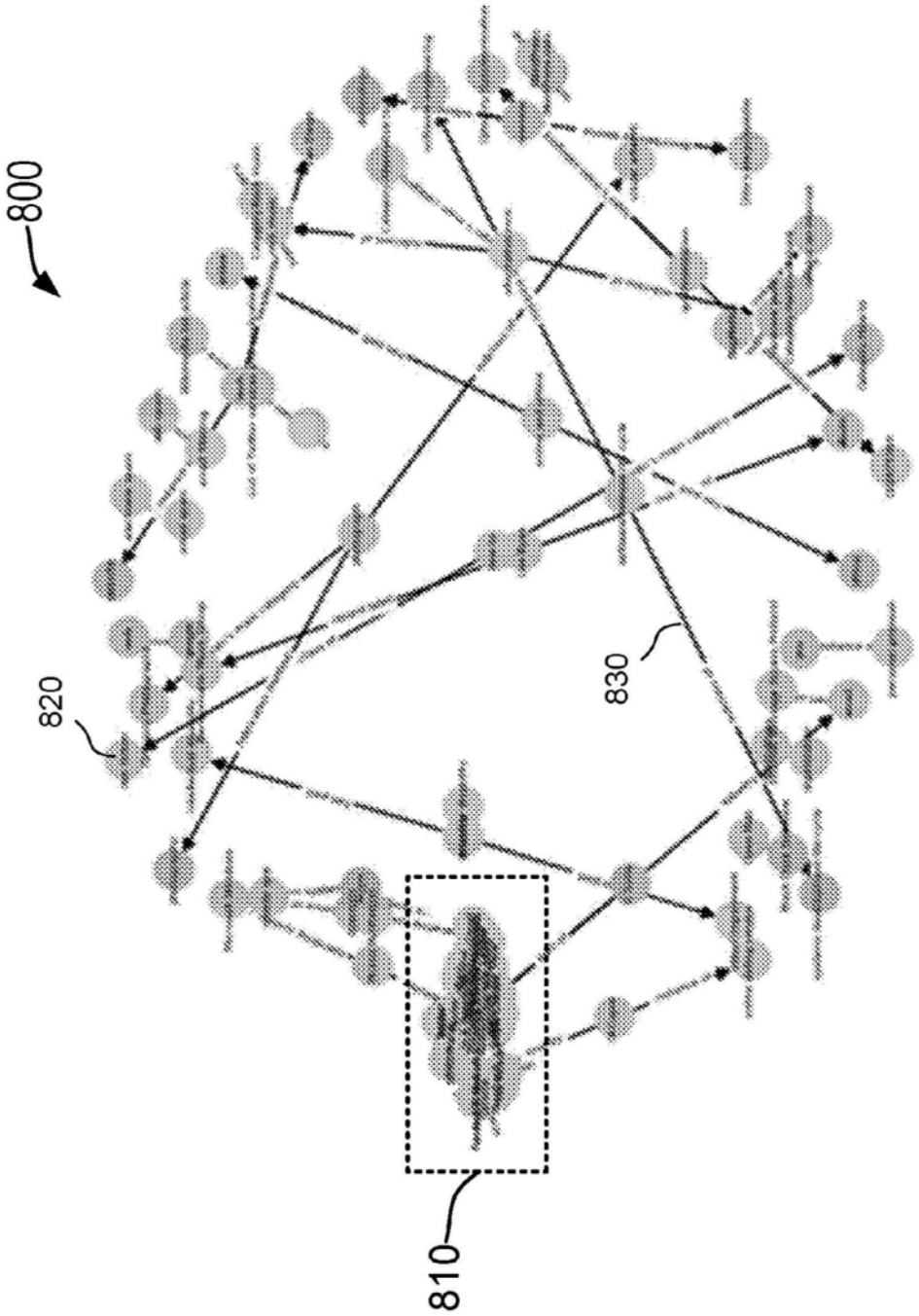


图8

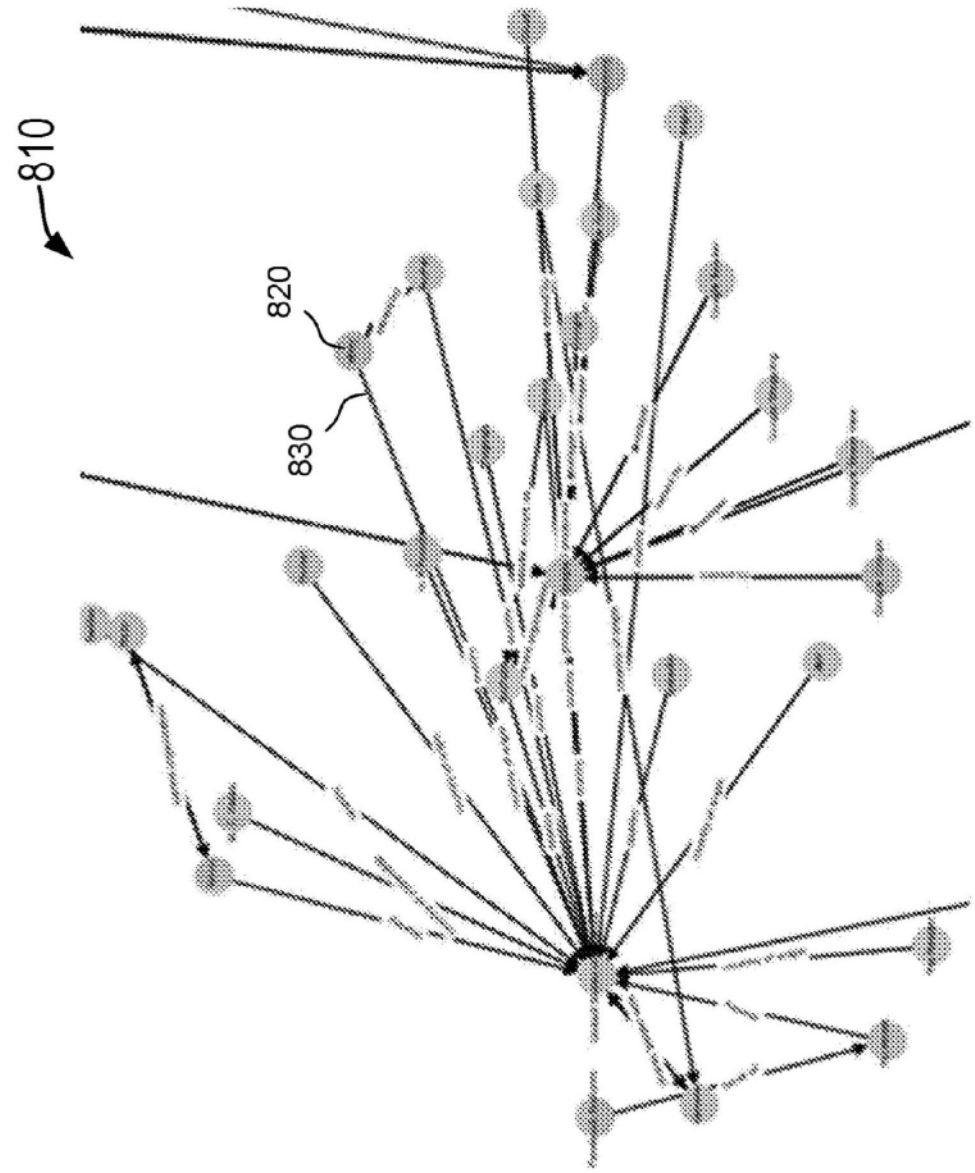


图9

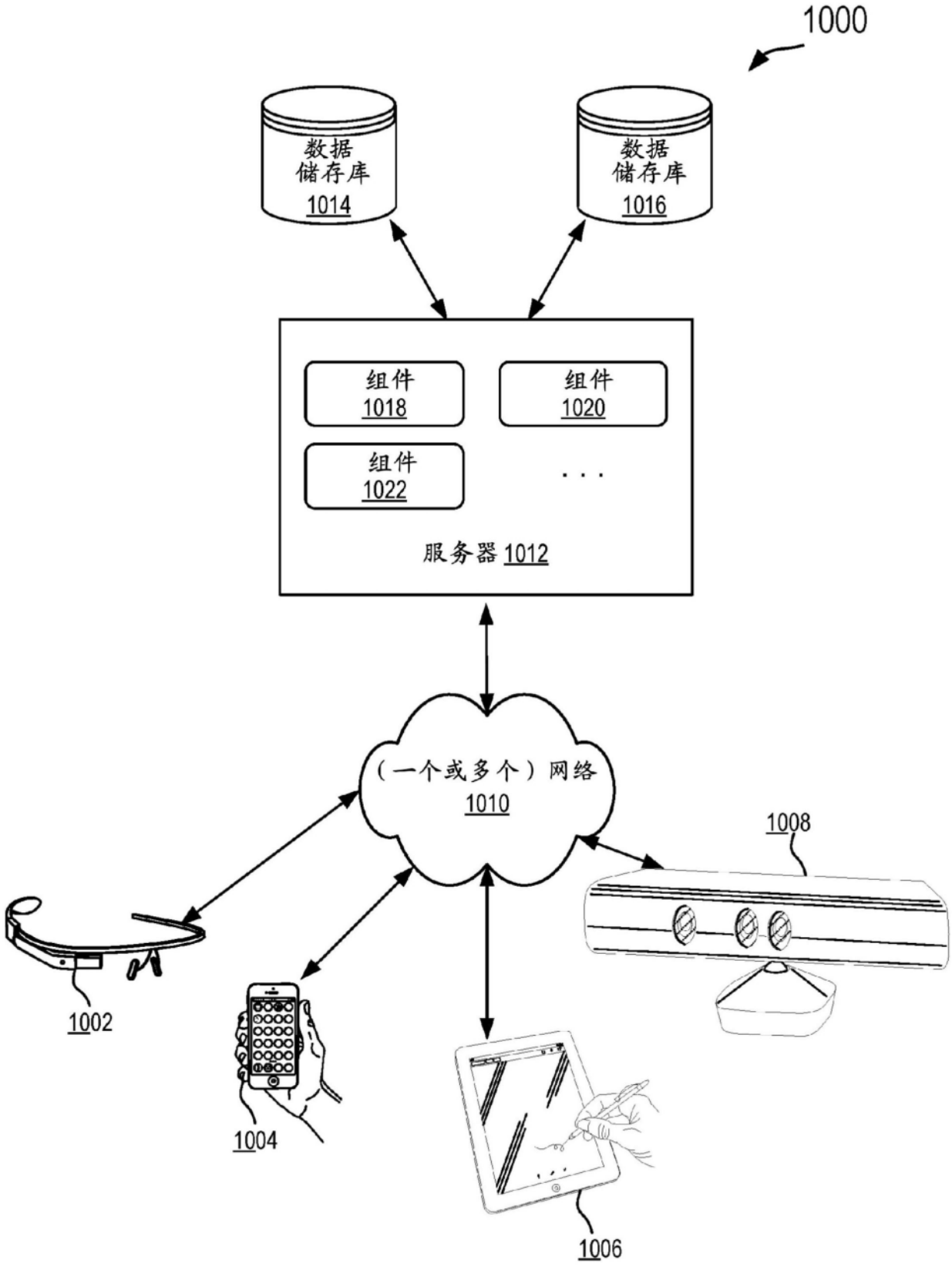

1000
数据储存库
1014
数据储存库
1016
组件
1018
组件
1020
组件
1022
. . .
服务器 1012
（一个或多个）网络
1010
1008
1002
1004
1006

图10

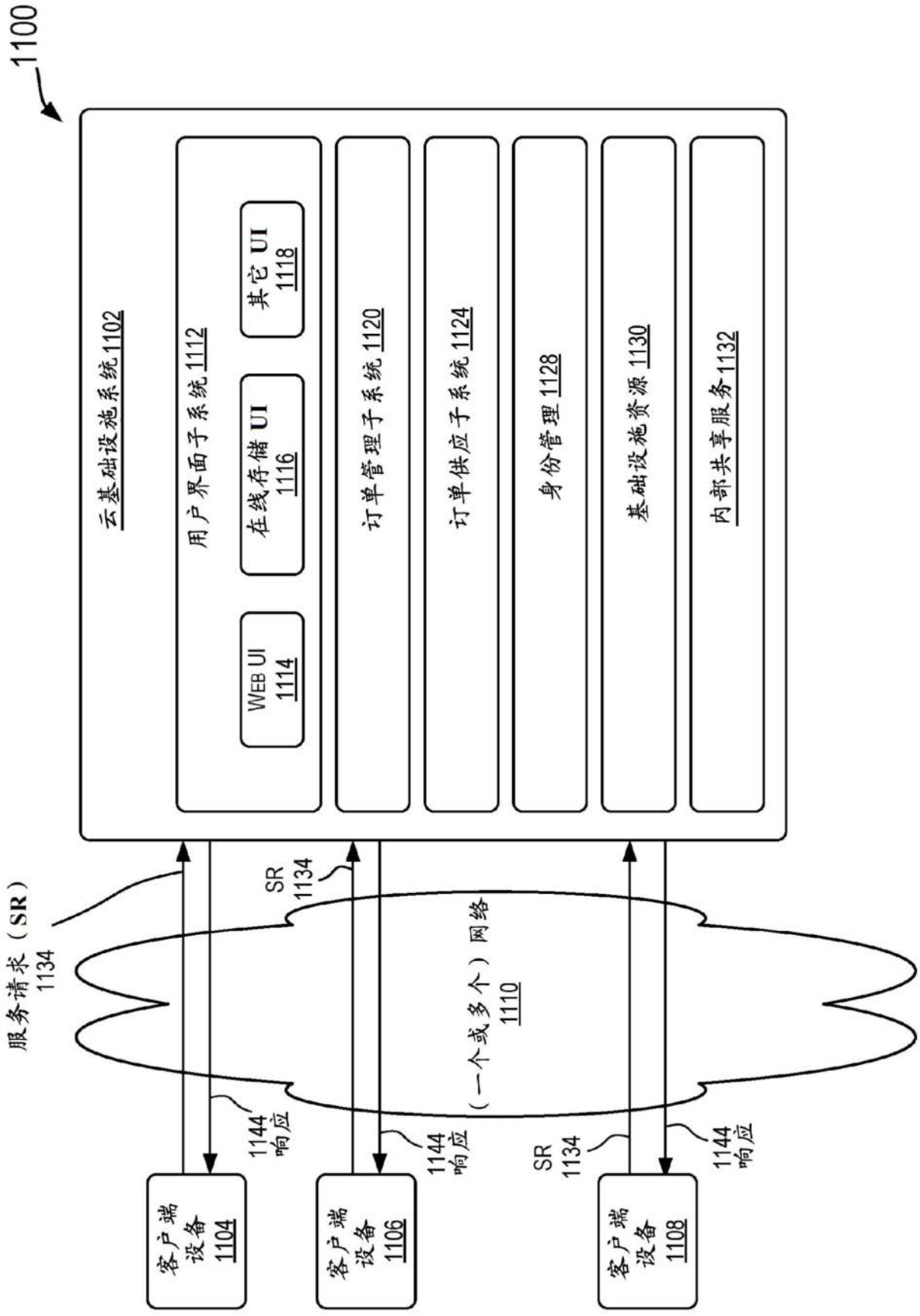

1100
云基础设施系统1102
用户界面子系统1112
WEB UI 1114
在线存储 UI 1116
其它 UI 1118
订单管理子系统1120
订单供应子系统1124
身份管理1128
基础设施资源1130
内部共享服务1132
服务请求（SR）1134
SR 1134
（一个或多个）网络 1110
1144 响应
1144 响应
SR 1134
1144 响应
客户端设备 1104
客户端设备 1106
客户端设备 1108

图11

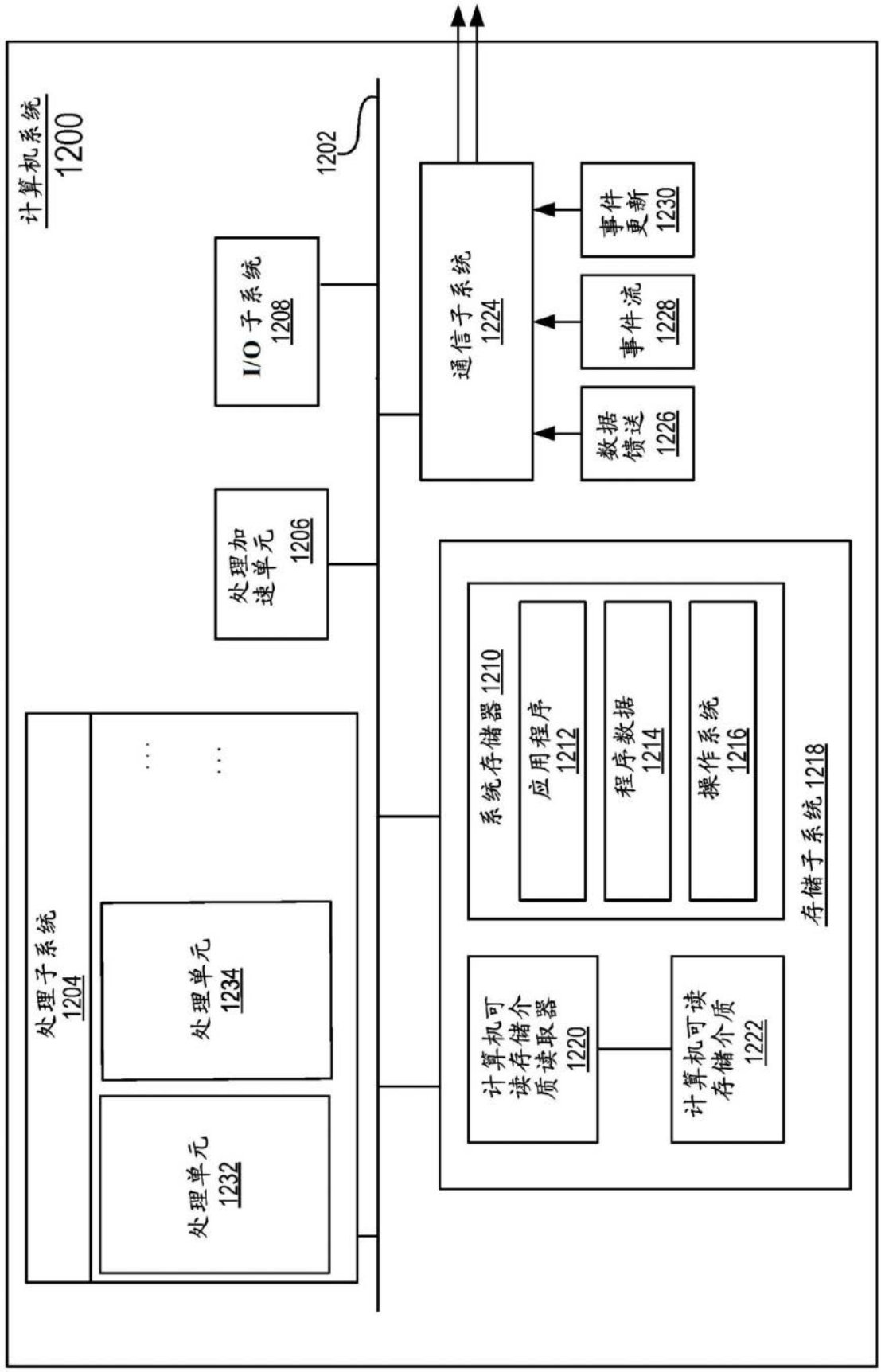

计算机系统
1200
1202
I/O 子系统
1208
处理加速单元
1206
通信子系统
1224
事件更新
1230
事件流
1228
数据馈送
1226
处理子系统
1204
处理单元
1232
处理单元
1234
系统存储器 1210
应用程序
1212
程序数据
1214
操作系统
1216
存储子系统 1218
计算机可读存储介质读取器
1220
计算机可读存储介质
1222

图12