



(12) 发明专利

(10) 授权公告号 CN 114375447 B

(45) 授权公告日 2025. 07. 08

(21) 申请号 202080063909.5

(22) 申请日 2020.09.08

(65) 同一申请的已公布的文献号  
申请公布号 CN 114375447 A

(43) 申请公布日 2022.04.19

(30) 优先权数据  
16/575,341 2019.09.18 US

(85) PCT国际申请进入国家阶段日  
2022.03.11

(86) PCT国际申请的申请数据  
PCT/IB2020/058338 2020.09.08

(87) PCT国际申请的公布数据  
W02021/053457 EN 2021.03.25

(73) 专利权人 国际商业机器公司  
地址 美国纽约阿芒克

(72) 发明人 O.西多尔金 S.巴丁

(74) 专利代理机构 北京市柳沈律师事务所  
11105  
专利代理师 陈金林

(51) Int.Cl.  
G06F 16/30 (2006.01)

(56) 对比文件  
CN 108363743 A, 2018.08.03  
CN 107885844 A, 2018.04.06

审查员 张晶

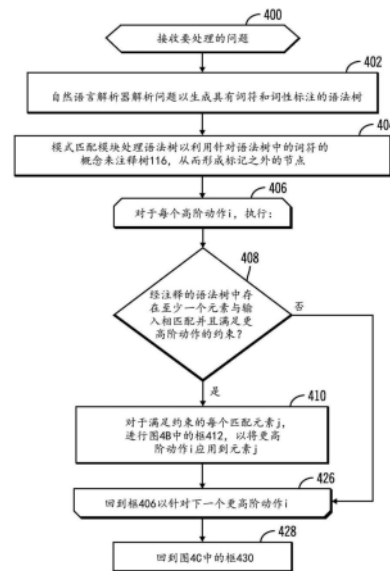
权利要求书5页 说明书10页 附图9页

(54) 发明名称

计算系统中的语言语句处理

(57) 摘要

一种使用更高阶动作来注释语法树的方法，包括：利用与语法树中的词符相对应的概念来注释语法树的信息空间。该概念包括问题的领域中的对象并且与领域的本体中的其他概念相关。更高阶动作指定输入、动作参数和输出。确定信息空间中的与更高阶动作的输入相对应的元素。确定问题的领域中的具有与所述动作参数相匹配的参数的动作。确定元素被输入到所确定的动作以产生输出。语法树的信息空间利用来自更高阶动作的输出被注释，以用于提供问题的答案。



1. 一种用于在计算系统中处理语言语句的计算机程序产品,其中所述计算机程序产品包括具有计算机可读程序指令的计算机可读存储介质,所述计算机可读程序指令由处理器执行以执行操作,所述操作包括:

根据要处理的问题生成语法树;

利用与所述语法树中的词符相对应的概念来注释所述语法树的信息空间,其中,所述概念包括所述问题的领域中的对象并且与所述领域的本体中的其他概念相关;

提供更高阶动作,所述更高阶动作指定输入、动作参数和输出;

确定所述语法树的所述信息空间中的与所述更高阶动作的输入相对应的元素;

确定所述问题的领域中的多个动作中的具有与所述动作参数相匹配的参数的动作;

提供所确定的元素作为所确定的动作的输入,以产生所确定的动作的输出,所确定的动作的输出根据由所述更高阶动作指定的输出来处理;以及

利用来自所述更高阶动作的输出来注释所述语法树的所述信息空间以用于提供所述问题的答案。

2. 根据权利要求1所述的计算机程序产品,其中,所述多个动作中的每一者具有参数:约束、输入和输出,所述输入包括所述约束所应用的所述语法树的所述信息空间中的元素,所述输出是所述动作响应于所述输入满足所述约束而产生的,其中,确定与所述动作参数相匹配的动作包括确定具有与所述更高阶动作中的所述动作参数的约束、输入和输出相匹配的约束、输入和输出的动作。

3. 根据权利要求1所述的计算机程序产品,其中,所述更高阶动作指定针对所述输入的约束,其中,确定所述语法树的所述信息空间中的与所述更高阶动作的输入相对应的元素包括:

确定所述语法树的信息空间是否包括满足所述更高阶动作的输入的约束的概念,其中,所确定的动作被应用于满足所述更高阶动作的输入的约束的概念,并且其中,来自所述更高阶动作的输出提供针对与所述更高阶动作的输入相对应的概念的真实数据。

4. 根据权利要求1所述的计算机程序产品,其中,确定所述语法树的所述信息空间中的满足所述更高阶动作的输入的元素包括确定所述语法树的信息空间中的多个元素,其中,所确定的动作被应用于所确定的多个元素中的每一个以产生所述更高阶动作的输出,从而注释所述语法树的信息空间的与所述更高阶动作的输入相对应的所确定的多个元素中的每一个。

5. 根据权利要求1所述的计算机程序产品,其中,所述更高阶动作和所述动作接收所述信息空间中的概念作为输入以产生包括针对所述概念的数据的输出,并且其中,所述操作还包括:

根据来自至少一个更高阶动作的针对概念而产生的数据来确定候选答案;

确定问题的每个候选答案的感官度量;以及

使用候选答案的感官度量来选择候选答案作为所述问题的答案。

6. 根据权利要求1所述的计算机程序产品,其中,所述更高阶动作的输入指示概念或数据,而不指定对所指示的概念或数据的约束,其中,所述操作还包括:

确定所述语法树的所述信息空间是否包括以下元素:所述元素包括在所述更高阶动作的输入中指示的概念或数据,其中,具有与所述动作参数相匹配的参数的动作响应于所述

语法树的所述信息空间中的与所述输入中指定的数据或概念相对应的元素而被处理,以产生所述更高阶动作的输出。

7. 根据权利要求1所述的计算机程序产品,其中,确定所述动作包括确定具有与所述动作参数相匹配的参数的多个动作,其中,所述操作还包括:

对于所确定的多个动作中的每一个,确定匹配强度,所述匹配强度指示所确定的动作的参数与所述动作参数的匹配的强度;以及

选择具有最高匹配强度的动作以应用于所述语法树的所述信息空间中的与所述更高阶动作的输入相对应的元素。

8. 根据权利要求1所述的计算机程序产品,其中,所述更高阶动作的输入指定概念以及对所述概念的约束,所述约束形成所述概念的集合,其中,根据所述动作参数确定的动作被应用于所述概念的集合。

9. 一种用于在计算系统中处理语言语句的计算机程序产品,其中,所述计算机程序产品包括计算机可读存储介质,所述计算机可读存储介质包括:

根据要处理的问题生成语法树,所述语法树包括信息空间,所述信息空间利用与所述语法树中的词符相对应的概念来注释,其中,所述概念包括所述问题的领域中的对象并且与所述领域的本体中的其他概念相关;

所述问题的领域中的具有参数的多个动作;以及

更高阶动作,其指定输入、动作参数和输出,其中,所述更高阶动作被处理以确定所述语法树的所述信息空间中的与所述更高阶动作的输入相对应的元素,并且确定所述多个动作中的具有与所述动作参数相匹配的参数的动作,其中,所确定的元素被提供作为所确定的动作的输入以产生所确定的动作的输出,所述输出根据由所述更高阶动作指定的输出来处理,并且其中,所述语法树的信息空间还利用来自所述更高阶动作的输出被进一步注释,以用于提供所述问题的答案。

10. 根据权利要求9所述的计算机程序产品,其中,所述多个动作中的每一者具有参数:约束、输入和输出,所述输入包括所述约束所应用的所述语法树的信息空间中的元素,所述输出是所述动作响应于所述输入满足所述约束而产生的,其中,具有与所述动作参数相匹配的参数的动作包括具有与所述更高阶动作中的所述动作参数的约束、输入和输出相匹配的约束、输入和输出的动作。

11. 根据权利要求9所述的计算机程序产品,其中,所述更高阶动作指定针对所述输入的约束,其中,所述语法树的所述信息空间中的与所述更高阶动作的输入相对应的所确定的元素包括满足所述更高阶动作的输入的约束的概念,其中,所确定的动作被应用于满足所述更高阶动作的输入的约束的概念,并且其中,来自所述更高阶动作的输出提供针对与所述更高阶动作的输入相对应的概念的真实数据。

12. 一种用于处理语言语句的系统,包括:

处理器;以及

计算机可读存储介质,其具有计算机可读程序指令,所述计算机可读程序指令由所述处理器执行以执行操作,所述操作包括:

根据要处理的问题生成语法树;

利用与所述语法树中的词符相对应的概念来注释所述语法树的信息空间,其中,所述

概念包括所述问题的领域中的对象并且与所述领域的本体中的其他概念相关；  
提供更高阶动作,所述更高阶动作指定输入、动作参数和输出；  
确定所述语法树的信息空间中的与所述更高阶动作的输入相对应的元素；  
确定所述问题的领域中的多个动作中的具有与所述动作参数相匹配的参数的动作；  
提供所确定的元素作为所确定的动作的输入以产生所确定的动作的输出,所确定的动作的输出根据由所述更高阶动作指定的输出来处理;以及  
利用来自所述更高阶动作的输出来注释所述语法树的所述信息空间以用于提供所述问题的答案。

13. 根据权利要求12所述的系统,其中,所述多个动作中的每一个具有参数:约束、输入和输出,所述输入包括所述约束所应用的所述语法树的信息空间中的元素,所述输出是所述动作响应于所述输入满足所述约束而产生的,其中,确定与所述动作参数相匹配的动作包括确定具有与所述更高阶动作中的所述动作参数的约束、输入和输出相匹配的约束、输入和输出的动作。

14. 根据权利要求12所述的系统,其中,所述更高阶动作指定针对所述输入的约束,其中,确定所述语法树的信息空间中的与所述更高阶动作的输入相对应的元素包括:  
确定所述语法树的信息空间是否包括满足所述更高阶动作的输入的约束的概念,其中,所确定的动作被应用于满足所述更高阶动作的输入的约束的概念,并且其中,来自所述更高阶动作的输出提供针对与所述更高阶动作的输入相对应的概念的真实数据。

15. 根据权利要求12所述的系统,其中,所述更高阶动作和所述动作接收所述信息空间中的概念作为输入以产生包括针对所述概念的数据的输出,并且其中,所述操作还包括:  
根据来自至少一个更高阶动作的针对概念而产生的数据来确定候选答案;  
确定所述问题的每个候选答案的感官度量;以及  
使用候选答案的感官度量来选择候选答案作为所述问题的答案。

16. 根据权利要求12所述的系统,其中,所述更高阶动作的输入指示概念或数据,而不指定对所指示的概念或数据的约束,其中,所述操作还包括:  
确定所述语法树的所述信息空间是否包括以下元素:所述元素包括所述更高阶动作的所述输入中指示的所述概念或数据,其中,具有与所述动作参数相匹配的参数的所述动作响应于所述语法树的所述信息空间中的与所述输入中指定的所述数据或概念相对应的元素而被处理,以产生所述更高阶动作的输出。

17. 根据权利要求12所述的系统,其中,确定所述动作包括确定具有与所述动作参数相匹配的参数的多个动作,其中,所述操作还包括:  
对于所确定的多个动作中的每一个,确定匹配强度,所述匹配强度指示所确定的动作的参数与所述动作参数的匹配的强度;以及  
选择具有最高匹配强度的动作以应用于所述语法树的信息空间中的与所述更高阶动作的输入相对应的元素。

18. 一种用于处理语言语句的系统,包括:  
处理器;以及  
具有由处理器执行的计算机可读程序代码的计算机可读存储介质,包括:  
根据要处理的问题生成语法树,所述语法树包括信息空间,所述信息空间利用与所述

语法树中的词符相对应的概念来注释,其中,所述概念包括所述问题的领域中的对象并且与所述领域的本体中的其他概念相关;

所述问题的领域中的具有参数的多个动作;以及

更高阶动作,其指定输入、动作参数和输出,其中,所述更高阶动作被处理以确定所述语法树的信息空间中的与所述更高阶动作的输入相对应的元素,并且确定所述多个动作中具有与所述动作参数相匹配的参数的动作,其中,所确定的元素被提供作为所确定的动作的输入以产生所确定的动作的输出,所确定的动作的输出根据由所述更高阶动作指定的输出来处理,并且其中,所述语法树的信息空间还利用来自所述更高阶动作的输出被进一步注释,以用于提供所述问题的答案。

19. 根据权利要求18所述的系统,其中,所述更高阶动作指定针对所述输入的约束,其中,所述语法树的信息空间中的与所述更高阶动作的输入相对应的所确定的元素包括满足所述更高阶动作的输入的约束的概念,其中,所确定的动作被应用于满足所述更高阶动作的输入的约束的概念,并且其中,来自所述更高阶动作的输出提供针对与所述更高阶动作的输入相对应的概念的真实数据。

20. 一种用于处理语言语句的方法,包括:

根据要处理的问题生成语法树;

利用与所述语法树中的词符相对应的概念来注释所述语法树的信息空间,其中,所述概念包括所述问题的领域中的对象并且与所述领域的本体中的其他概念相关;

提供更高阶动作,所述更高阶动作指定输入、动作参数和输出;

确定所述语法树的信息空间中的与所述更高阶动作的输入相对应的元素;

确定所述问题的领域中的多个动作中的具有与所述动作参数相匹配的参数的动作;

提供所确定的元素作为所确定的动作的输入以产生所确定的动作的输出,所确定的动作的输出根据由所述更高阶动作指定的输出来处理;以及

利用来自所述更高阶动作的输出注释所述语法树的所述信息空间,以用于提供所述问题的答案。

21. 根据权利要求20所述的方法,其中,所述多个动作中的每一个具有参数:约束、输入和输出,所述输入包括所述约束所应用的所述语法树的所述信息空间中的元素、以及所述输出是所述动作响应于所述输入满足所述约束而产生的,其中,确定与所述动作参数相匹配的动作包括确定具有与所述更高阶动作中的所述动作参数的约束、输入和输出相匹配的约束、输入和输出的动作。

22. 根据权利要求20所述的方法,其中,所述更高阶动作指定针对输入的约束,其中,确定所述语法树的所述信息空间中的与所述更高阶动作的输入相对应的元素包括:

确定所述语法树的信息空间是否包括满足所述更高阶动作的所述输入的所述约束的概念,其中,所确定的动作被应用于满足所述更高阶动作的输入的约束的概念,并且其中,来自所述更高阶动作的输出提供针对与所述更高阶动作的输入相对应的概念的真实数据。

23. 根据权利要求20所述的方法,其中,所述更高阶动作和所述动作接收所述信息空间中的概念作为输入以产生包括针对所述概念的数据的输出,并且其中,所述方法还包括:

根据来自至少一个更高阶动作的针对概念而产生的数据来确定候选答案;

确定所述问题的每个候选答案的感官度量;以及

使用候选答案的感官度量来选择候选答案作为所述问题的答案。

24. 根据权利要求20所述的方法, 其中, 所述更高阶动作的输入指示概念或数据, 而不指定对所指示的概念或数据的约束, 其中, 所述方法还包括:

确定所述语法树的所述信息空间是否包括以下元素: 所述元素包括在所述更高阶动作的输入中指示的概念或数据, 其中, 具有与所述动作参数相匹配的参数的动作响应于所述语法树的所述信息空间中的与所述输入中指定的所述数据或概念相对应的元素而被处理, 以产生所述更高阶动作的输出。

25. 根据权利要求20所述的方法, 其中, 确定所述动作包括确定具有与所述动作参数相匹配的参数的多个动作, 其中, 所述方法还包括:

对于所确定的多个动作中的每一个, 确定匹配强度, 所述匹配强度指示所确定的动作的参数与所述动作参数的匹配的强度; 以及

选择具有最高匹配强度的动作以应用于所述语法树的信息空间中与所述更高阶动作的输入相对应的元素。

## 计算系统中的语言语句处理

### 技术领域

[0001] 本发明总体上涉及计算机系统上的语言语句处理,并且更具体地涉及使用更高阶动作利用用于生成问题的答案的针对概念的真实数据来注释语法树的计算机程序产品、系统和方法。

### 背景技术

[0002] 来自国际商业机器公司的嵌入式商业人工智能框架允许客户针对领域特定的自然语言处理环境创建概念和属性的函数和本体,这些函数和本体用于辨识和推理自然语言模式,以执行自然语言理解和将语言标记链接到适当的概念。客户可以在领域中提供动作,该动作处理根据要理解的句子形成的语法树中的节点以生成针对概念的真实数据。经注释的概念用于理解句子或提供对在语法树中建模的问题的答案。在对问题进行推理的过程中,给定一组动作和规则,系统考虑它可以采取的所有可能的结果,最终在推理引擎不能找到答案的情况下选择最好的一个或请求用户答疑。

[0003] 在本领域中需要提供用于将真实数据添加到正被处理的句子的信息空间中以改进自然语言处理的改进技术。

### 发明内容

[0004] 提供了一种使用更高阶动作利用用于生成问题的答案的针对概念的真实数据来注释语法树的计算机程序产品、系统和方法。语法树是根据要处理的问题生成的。语法树的信息空间利用与语法树中的词符相对应的概念来注释,其中概念包括问题的领域中的对象并且与领域的本体中的其他概念相关。更高阶动作指定输入、动作参数和输出。确定语法树的信息空间中的与更高阶动作的输入相对应的元素。确定问题的领域中的多个动作中的具有与动作参数相匹配的参数的动作。所确定的元素被提供作为所确定的动作的输入,以产生所确定的动作的输出,该输出根据由更高阶动作指定的输出来处理。语法树的信息空间利用来自更高阶动作的输出注释,以用于提供问题的答案。

[0005] 还提供了一种计算机程序产品和系统,其包括根据要处理的问题生成语法树,该语法树包括信息空间,该信息空间利用与语法树中的词符相对应的概念来注释。概念包括问题的领域中的对象并且与领域的本体中的其他概念相关;问题的领域中的具有参数的多个动作。更高阶动作指定输入、动作参数和输出。更高阶动作被处理以确定语法树的信息空间中的与更高阶动作的输入相对应的元素,并且确定多个动作中的具有与动作参数相匹配的参数的动作。所确定的元素被提供作为所确定的动作的输入,以产生所确定的动作的输出,该输出根据由更高阶动作指定的输出来处理。语法树的信息空间利用来自更高阶动作的输出被进一步注释,以用于提供问题的答案。

[0006] 上述实施例通过提供更高阶动作提供对用于注释语法树以进行自然语言处理的计算机技术的改进,该更高阶动作采取动作参数作为参数以与要调用以处理输入的现有动作相匹配。更高阶动作被处理以确定具有与为该更高阶动作指定的动作参数相匹配的参数

的所有动作。这允许开发者定义指定动作参数或签名的更高阶动作,以用于确定其他先前定义的动作,以从语法树中的当前元素s产生输出,以用于注释语法树的信息空间。

[0007] 实施例的主题可以可选地包括可选实施例,其中多个动作中的每一个具有参数:约束、输入以及输出,该输入包括约束所应用的语法树的信息空间中的元素,该输出是响应于满足约束的输入而由动作产生的。确定与动作参数相匹配的动作包括确定具有与更高阶动作中的动作参数的约束、输入和输出相匹配的约束、输入和输出的动作。

[0008] 在上述实施例的情况下,对输入的约束允许用户指定针对输入参数的约束或条件,以确保更高阶动作适用于满足约束条件的那些输入,以提供对更高阶动作的适用性的更细粒度的控制。这样,具有与更高阶动作的约束、输入和输出参数相匹配的约束、输入和输出参数的动作被确定。

[0009] 实施例的主题可以可选地包括可选实施例,其中更高阶动作指定针对输入的约束。确定语法树的信息空间中与更高阶动作的输入相对应的元素包括确定语法树的信息空间是否包括满足更高阶动作的输入的约束的概念。所确定的动作被应用于满足更高阶动作的输入的约束的概念。来自更高阶动作的输出提供了针对与更高阶动作的输入相对应的概念的真实数据。

[0010] 在上述实施例的情况下,更高阶动作被用来确定满足针对输入的约束的概念,使得所确定的动作可以被应用于满足该约束的概念,以提供针对该概念的数据的输出,以在语法树的信息空间中进行注释。这允许更高阶动作触发其他限定动作的应用以应用于信息空间中的概念,从而生成针对信息空间中的概念的数据,以改进对语法树中表示的问题的答案的确定。

[0011] 实施例的主题可以可选地包括可选实施例,其中更高阶动作和动作接收信息空间中的概念作为输入以产生包括针对概念的数据的输出。根据来自至少一个更高阶动作的针对概念的数据来确定问题的候选答案以及每一个候选答案的感官度量。候选答案的感官度量用于选择候选答案作为问题的答案。

[0012] 在上述实施例的情况下,如果存在通过处理来自至少一个更高阶动作的针对概念产生的数据而得到的多个候选答案,则候选答案的感官度量被用来选择具有最佳感官度量值的候选答案以返回问题的最佳答案。

[0013] 实施例的主题可以可选地包括可选实施例,其中确定动作包括确定具有与动作参数相匹配的参数的多个动作。对于所确定的多个动作中的每一个,确定匹配强度,该匹配强度指示所确定的动作的参数与动作参数的匹配的强度。选择具有最高匹配强度的动作以应用于语法树的信息空间中的与更高阶动作的输入相对应的元素。

[0014] 在上述实施例的情况下,如果存在多个具有与更高阶动作的动作参数相匹配的参数的所确定的动作,则确定动作的匹配强度被以选择动作来处理信息空间中具有最高匹配强度的元素。这允许确定与更高阶动作参数相匹配的多个不同的可能动作,以扩展被认为是匹配的动作的数量,从而增加找到满足更高阶动作参数的最佳动作的可能性。

## 附图说明

[0015] 图1示出了自然语言计算系统的实施例。

[0016] 图2示出了用于利用真实数据注释语法树信息空间的动作的实施例。

[0017] 图3示出了更高阶动作的实施例,该更高阶动作使用领域中已经存在的动作,将动作作为参数,利用真实数据进一步注释语法树信息空间。

[0018] 图4a、图4b和图4c示出了用于生成问题的信息空间以用于生成问题的答案的操作的实施例。

[0019] 图5示出了根据问题术语生成的语法树的示例。

[0020] 图6示出了利用语法树中的问题词符(token)的概念被进一步注释的语法树的示例。

[0021] 图7示出了图6的经注释的语法树,其被表示为针对问题词符的节点和链接到问题词符的概念。

[0022] 图8示出了通过处理应用于经注释的语法树中的概念和数据节点的动作和更高阶动作,利用链接到概念的真实数据被进一步注释的语法树的示例。

[0023] 图9描绘了其中可以实施图1的组件的计算环境。

### 具体实施方式

[0024] 在当前系统中,用户可以在领域中创建动作,这些动作用于针对语法树中的概念,将真实数据生成到语法树的节点中。动作被编码为基于经注释的语法树中的概念和数据来执行特定的数据收集操作,诸如确定满足属性的多个项、项的数量或描述,以及从数据库或用户收集其他信息。由动作输出的数据可以被添加到信息空间,并且用于回答在信息空间中的语法树中建模的问题。

[0025] 所描述的实施例通过提供更高阶动作来提供针对用于对语法树的信息进行自然语言处理的计算机技术的改进,该更高阶动作采取动作参数作为参数来与要调用的现有动作相匹配,以处理可以被定义为项或概念的集合的输入。该更高阶动作被处理以确定具有与特定于更高阶动作的动作参数相匹配的参数的所有动作。这允许开发者定义指定动作参数或签名的更高阶动作,以用于确定领域中的具有在某种置信度上与在更高阶动作中特定的动作参数相匹配的参数或签名的其它先前定义的动作。然后,可以将领域中的与更高阶动作的参数相匹配的所确定的先前定义的动作应用于经注释的语法树中的概念或数据,包括经注释的语法树中的项的集合。

[0026] 图1示出了具有用于执行存储在存储器/存储装置104中的程序组件的处理器102的计算机系统的实施例。存储器/存储装置104包括由处理器102执行的程序组件,包括自然语言解析器106、模式匹配模块108、语义动作模块110和机器推理模块112。自然语言解析器106将单词的问题114解析到语法树116中,包括语法树116的初始信息空间,诸如通过执行词性标注来基于单词的定义、上下文和词形还原将其标记为对应于特定的词性,以基于其含义确定单词的词性;模式匹配模块108,用于将来自问题114的语法树118中的自然语言词符与领域中的概念进行匹配,在该领域中,问题被要求产生经注释的语法树118或语法树118的更新的信息空间,其中,概念提供关于问题114中的词符/语言的含义的属性和信息。

[0027] 由模式匹配模块108捕获的模式添加语法上下文,在该语法上下文中,问题114的词符被使用。例如,术语“weather”在这种情况下表示名词主语。然而,单词“weather”可以在其他句子中包含非常不同的语义含义,例如“can we weather the storm?”。基于上述自然语言模式,模式匹配模块112理解“weather”的这种使用不同于第一使用,并且随后,在回

答第二类型的问题时,它将不会将概念“:Weather”引入其推理流水线。在术语之前的冒号,例如“:Term”表示“Term”是在问题114被问到的领域的本体中定义的概念。概念可以包括领域的数据库模型中的属性。模式匹配模块108使得能够将问题114中的自然语言映射到领域的本体中的概念。针对问题114中的词符可以有概念树,其包括单独列出的注释,形成提供用于推理问题114中的术语的含义的起点的单独概念的注释树。注释语法树116中的词符的概念允许系统在处理问题114时将这些概念作为问题114的词符的一部分来考虑和辨识。例如,如果问题请求“show trending products”,则术语“trending”可以利用概念“:Trending”来注释,提供关于trending在特定领域中是什么含义的信息,在该特定领域中,关于产品的问题将被问到,并且术语“product”可以利用概念“Product”来注释,提供关于在领域中产品的含义的信息,诸如针对特定零售商。语法树116中的每个词符可以用概念来注释。

[0028] 语义动作模块110利用所声明的动作200和更高阶动作300来处理经注释的语法树118中的节点,该节点包括来自问题114的词符以及针对词符的所注释的概念,该更高阶动作包括输出与经注释的语法树118中的概念相对应的真实数据的函数。语义动作模块110的输出是利用概念和真实数据来注释的经进一步注释的语法树120或语法树120的经进一步更新的信息空间,其中,真实数据可以提供来自与概念相关的数据库的系统信息,诸如满足概念的产品列表、产品数量等。机器推理模块112处理具有针对概念的真实数据的经注释的语法树120,并产生问题114的候选答案122,该候选答案122具有感官度量,该感官度量指示答案122对问题114做出响应的值或强度。在一个实施例中,感官度量可以基于用于生成候选答案的单词的数量、为该答案处理的概念的数量、以及问题114和答案122的语义和语法表示之间的相似性。每个被应用的动作 $200_i$ 和更高阶动作 $300_i$ 可以递增地更新语法树120的信息空间,使得随后的动作 $200_i$ 和更高阶动作 $300_i$ 可以处理作为输入输出的、从先前执行的动作和更高阶动作添加到语法树120的信息空间的内容。

[0029] 存储器/存储装置104可以包括用于存储要执行的程序和由程序110用来执行的信息的适当的易失性或非易失性存储器。

[0030] 通常,诸如程序组件106到124之类的程序模块可以包括执行特定任务或实施特定抽象数据类型的例程、程序、对象、组件、逻辑、数据结构等。图1的计算机系统100的程序组件和硬件设备可以在一个或多个计算机系统中实施,其中如果它们在多个计算机系统中实施,则计算机系统可以通过网络通信。

[0031] 程序106、108、110、112、200、300可以包括加载到存储器中并由处理器执行的程序代码。或者,功能中的一些或全部可以在硬件设备中实施,例如在专用集成电路(ASIC)中实施,或者由单独的专用处理器执行。

[0032] 在一个实施例中,程序106、108、110、112可以实施诸如决策树学习、关联规则学习、神经网络、归纳编程逻辑、支持向量机、贝叶斯网络之类的机器学习技术,以在自然语言处理流水线中执行它们的特定任务。程序106、108、110、112可以包括人工神经网络程序。每个神经网络可以使用后向传播来训练,以调整隐藏层中的节点处的权重和偏差,从而产生计算的输出,诸如语法树116、具有概念的经注释的语法树118,执行动作以利用真实数据进一步注释语法树120、以及生成并回答问题。在用于训练神经网络机器学习模块的后向传播中,隐藏层中的节点处的偏差被相应地调整以基于指定的置信度水平来产生期望的结果。

后向传播可以包括使用梯度下降法来监督学习人工神经网络的算法。给定人工神经网络和误差函数,该方法可以相对于神经网络的权重和偏差来计算误差函数的梯度。

[0033] 在图1中,箭头被示出在存储器/存储装置104中的组件之间。这些箭头表示去往和来自程序组件106、108、110和112的信息流,并且不表示存储器104中的数据结构。

[0034] 图2示出了动作 $200_i$ 的实例的实施例,包括参数,也被称为签名,该参数包括约束202(其是可选的)、输入204(诸如,概念或数据)、输出206、以及用于将输入204变换成输出206的代码208。对于满足约束202的输入204,诸如经注释的语法树118中的概念或数据节点,代码208根据输入204生成输出206,诸如针对概念的真实数据节点,诸如由概念204标识的项的多个实例,以及关于概念的其他信息。

[0035] 约束202可以包括概念的限定条件,并且包括主语(输入204)、谓语和宾语,其中谓语将主语约束到宾语。例如,“a subclassOf.list”将符号a限制为任何列表,使得经注释的语法树118中的概念节点必须包括列表以便经受动作 $200_i$ 。输入204可以包括必须与要被选择执行的动作相匹配的概念树。输出206可以包括概念的平面列表或包括概念的真实数据。签名内的概念可以利用附加规范来进一步限定。输入的默认参数可以是概念。然而,输入还可以包括由另一动作针对概念产生的真实数据,诸如真实产品数据、真实发票数据等。

[0036] 动作的示例包括:

[0037] :ActionShow-概念,其表示用户对显示数据的请求。它将产生:可显示的数据。

[0038] :ActionModify-概念,其表示用户对更新数据的请求。它将修改数据并向用户发布消息。

[0039] :ActionDelete-概念,其表示用户对移除数据的请求。它将移除数据并产生对用户的文本响应。

[0040] 动作的另一示例是GetTrendingProducts,其具有签名“Products (Trending) -> data:products”。该动作指定对于热门的(其是约束)被输入的概念“Products”,将输出关于热门的产品的数据,诸如最近流行的产品,即热门的。

[0041] 通过名称(SearchProductByName)在数据库中搜索产品的动作的示例可以具有签名“Product (optional:WithName (data:UserString)) ->data Product”,使得对于树中的被约束为具有名称“UserString”的概念,将输出数据:Product,包括具有该产品名称的产品。

[0042] 图3示出了更高阶动作 $300_i$ 的实例的实施例,其包括针对输入304的约束302,包括注释语法树118中的节点、动作参数306(诸如,签名)、指示参数(诸如,约束、输入和输出),以用于寻找具有与动作签名306的参数相匹配的参数202、204、206的一个或多个匹配动作、以及与动作参数306相匹配的针对由动作 $200_i$ 产生的输出206的输出308。

[0043] 对于输入304,树节点被扫描以确定与输入304相匹配并且满足所提供的任何约束302的概念或数据,然后将满足约束302的概念或数据304作为输入204提供给具有与动作参数306相匹配的参数202、204、206的动作 $200_i$ 。更高阶动作 $300_i$ 还可以进一步具有用于将输入304变换成输出306的代码310。

[0044] 通过更高阶动作 $300_i$ ,领域动作的开发者可以将动作参数指定为306,以用于寻找相匹配的动作 $200_i$ 以应用于输入304的集合,诸如修改输入304的约束的集合,诸如包括或排除输入304。

[0045] 例如,具有排除项的更高阶的动作可以是以下形式:

[0046] `ExcladedeItems:xs isListOf a=>ActionExclude(data xs,action[data xs->data xs]) ->:Message`,其中,动作名称“ExcladedeItems”具有概念xs作为输入,该概念xs被约束为列表,其中一些项具有属性a。`ActionExclude`指定随后的动作参数306采用集合xs,包括列表a,并返回集合,然后在消息中输出该集合。

[0047] 这样,更高阶动作 $300_i$ 采取另一动作作为参数,并采取数据集合作为输入,包括针对输入304节点的约束302(包括概念或数据节点),然后对数据集合作执行与动作参数306相匹配的具有参数202、204的动作 $200_i$ ,以便包括或排除项,诸如基于由匹配的动作 $200_i$ 定义的属性、参数或查询来从由输入304所形成的数据集中定位项的子集。在图3的实施例的情况下,开发者或用户不必预先确定指定动作,但可定位具有与动作参数306的参数(例如,签名)相匹配的参数的动作,以用于基于输入304和约束302从集合中排除数据,并产生输出数据以包括在经注释的语法树120的信息空间中。如果更高阶动作 $300_i$ 指定代码310,则如果输入304满足约束302,该代码310可被进一步执行。

[0048] 图4a、图4b和图4c提供了由自然语言解析器106、模式匹配模块108、语义动作模块110和机器推理模块112执行的操作的实施例,以处理输入问题114,从而确定为问题114提供的答案124。在接收到(在框400)要处理的问题114时,自然语言解析器106解析(在框402)该问题以生成具有词符和词性标注的语法树116。模式匹配模块108处理语法树116,以利用针对语法树中的词符的概念来注释树116,从而形成词符之外的节点,以产生经注释的语法树118。语义动作模块110对问题114的领域中的每个更高阶动作 $300_i$ 执行框406至426处的操作的循环。如果(在框408)在树118中没有相匹配的元素,则控制进行到框426以处理下一个更高阶动作 $300_i$ ,直到所有更高阶动作300都被处理。如果(在框408)在注释语法树118中存在与更高阶动作 $300_i$ 的输入304相匹配并且满足约束302的至少一个元素,例如概念或数据,则对于满足约束302的每个相匹配的元素j,控制继续(在框410)以执行图4b中的框412到426,以将更高阶动作i应用于元素j。

[0049] 如果(在框412)更高阶动作 $300_i$ 中的动作参数306与针对领域的动作 $200_k$ 中的至少一个的参数(约束202、输入204、输出206)相匹配,并且如果(在框414)存在具有多个匹配参数的动作 $200_i$ ,则语义动作模块110选择(在框416)具有匹配参数202、204、206的动作中的具有与问题114最匹配的语法的一个动作。否则,如果(在框414)仅存在一个匹配动作 $200_k$ ,则选择该一个匹配动作 $200_k$ (在框418)。从框416或418,(在框420)针对所选择的动作 $200_k$ 的代码208被执行,以产生针对所确定的元素j的输出206,例如,真实数据,诸如多个项或满足准则的多个项。来自所执行的所选择的动作 $200_k$ 的输出206根据更高阶动作i的输出308被输出(在框422处),并且作为数据或某些其他元素被添加到经注释的语法树120。如果没有匹配动作或在框422处输出实际数据之后,根据框412的否分支,控制继续(在框424)以返回到图4a中的框410。此外,如果被提供了,则也可以执行针对更高阶动作 $300_i$ 的代码310。

[0050] 在处理了所有的更高阶动作300以将真实数据添加到经注释的语法树120之后,控制(在框428)继续进行到图4c中的框430,以继续处理经注释的语法树120,以确定最终答案124。在框430,语义动作模块110可以进一步处理动作200,以输出针对经注释的语法树120中的输入节点204的真实数据,从而为信息空间提供进一步的真实数据来回答问题114,诸如提供来自数据库的信息或针对概念的其他特定信息,例如满足标准的多个产品、产品列

表等。机器推理模块112处理经注释的语法树120,以根据通过处理动作200和更高级动作300被添加到经注释的语法树的真实数据确定(在框432)所接收的问题的候选答案122。如果(在框434)仅存在一个候选答案122,则返回(在框436)该候选答案122作为问题114的最终答案124。如果(在框434)存在多个候选答案122,则针对每个候选答案122确定(在框438)感官度量,该感官度量指示该候选答案122是问题114的响应的程度。返回(在框440)具有最佳感官度量的候选应答122作为最终应答124。

[0051] 在图4a、图4b、图4c的实施例的情况下,通过使更高阶动作采取具有与特定于更高阶动作的动作参数相匹配的约束输入和输出的另一动作作为输入,更高阶动作可用于排除或修改对其应用了动作的集合中的项,以找到可以执行由更高阶动作300<sub>i</sub>中的动作参数指定的操作的所有动作。所描述的实施例允许重新使用现有的语义动作,以便通过指定针对概念的项的集合或列表并然后指定动作的签名或参数来寻找要应用于项的集合的一个或多个动作,来开发新的功能。

[0052] 图5示出了针对示例问题“show me a list of all contacts in my organization”,利用问题中的词符和词符的函数,从自然语言解析器106中产生的语法树500(诸如,语法树116)的初始信息空间的示例。

[0053] 图6示出了利用概念注释的经注释的语法树600(诸如,注释语法树118)的信息空间的示例,此概念诸如,提供针对语法树116的项或词符的概念,:ActionShow、:ContactLists、:Contains delation、Organization和:Own。

[0054] 图7示出了以节点形式表示语法树600的语法树700的信息空间,其中具有问题114中的术语的节点被示出为实心填充,并链接到示出为具有白色中心的概念节点。

[0055] 图8示出了在应用动作200以添加链接到概念节点的真实数据节点之后的语法树800的经进一步注释的信息空间,诸如经注释的语法树120,其中由动作生成的真实数据节点被示为具有彩色中心的白色节点。变体面板802示出了用于生成真实数据节点的动作,真实数据节点被示出为具有实心中心的白色节点,概念被示出为具有白色中心的实心节点。经注释的语法树800可被处理以通过考虑针对链接到问题词符或术语的概念而生成的真实数据来生成问题的候选答案。

[0056] 本发明可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括其上具有计算机可读程序指令的计算机可读存储介质(或多个介质),所述计算机可读程序指令用于使处理器执行本发明的各方面。

[0057] 计算机可读存储介质可以是能够保留和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质可以是例如但不限于电子存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或前述的任何合适的组合。计算机可读存储介质的更具体示例的非穷举列表包括以下:便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式光盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、诸如上面记录有指令的打孔卡或凹槽中的凸起结构的机械编码装置,以及上述的任何适当组合。如本文所使用的计算机可读存储介质不应被解释为暂时性信号本身,诸如无线电波或其他自由传播的电磁波、通过波导或其他传输介质传播的电磁波(例如,通过光纤线缆的光脉冲)、或通过导线传输的信号。

[0058] 本文描述的计算机可读程序指令可以从计算机可读存储介质下载到相应的计算/处理设备,或者经由网络,例如因特网、局域网、广域网和/或无线网络,下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光传输光纤、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或网络接口从网络接收计算机可读程序指令,并转发计算机可读程序指令以存储在相应计算/处理设备内的计算机可读存储介质中。

[0059] 用于执行本发明的操作的计算机可读程序指令可以是汇编指令、指令集架构 (ISA) 指令、机器相关指令、微代码、固件指令、状态设置数据,或者以一种或多种编程语言的任意组合编写的源代码或目标代码,所述编程语言包括面向对象的编程语言,例如Java、Smalltalk、C++等,以及常规的过程式编程语言,例如“C”编程语言或类似的编程语言。计算机可读程序指令可以完全在用户的计算机上执行,部分在用户的计算机上执行,作为独立的软件包执行,部分在用户的计算机上并且部分在远程计算机上执行,或者完全在远程计算机或服务器上执行。在后一种情况下,远程计算机可以通过任何类型的网络连接到用户的计算机,包括局域网 (LAN) 或广域网 (WAN), 或者可以连接到外部计算机 (例如,使用因特网服务提供商通过因特网)。在一些实施例中,为了执行本发明的各方面,包括例如可编程逻辑电路、现场可编程门阵列 (FPGA) 或可编程逻辑阵列 (PLA) 的电子电路可以通过利用计算机可读程序指令的状态信息来执行计算机可读程序指令以使电子电路个性化。

[0060] 在此参考根据本发明实施例的方法、装置 (系统) 和计算机程序产品的流程图和/或框图来描述本发明的各方面。将理解,流程图和/或框图的每个框以及流程图和/或框图中的框的组合可以由计算机可读程序指令来实现。

[0061] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器以产生机器,使得经由计算机或其他可编程数据处理装置的处理器执行的指令创建用于实现流程图和/或框图的一个或多个框中指定的功能/动作的装置。这些计算机可读程序指令还可以存储在计算机可读存储介质中,其可以引导计算机、可编程数据处理装置和/或其他设备以特定方式工作,使得其中存储有指令的计算机可读存储介质包括制品,该制品包括实现流程图和/或框图的一个或多个框中指定的功能/动作的各方面的指令。

[0062] 计算机可读程序指令还可以被加载到计算机、其他可编程数据处理装置或其他设备上,以使得在计算机、其他可编程装置或其他设备上执行一系列操作步骤,以产生计算机实现的过程,使得在计算机、其他可编程装置或其他设备上执行的指令实现流程图和/或框图的一个或多个框中指定的功能/动作。

[0063] 附图中的流程图和框图示出了根据本发明的各种实施例的系统、方法和计算机程序产品的可能实现的架构、功能和操作。在这点上,流程图或框图中的每个框可以表示指令的模块、段或部分,其包括用于实现指定的逻辑功能的一个或多个可执行指令。在一些替代实施方案中,框中所提及的功能可不按图中所提及的顺序发生。例如,连续示出的两个框实际上可以基本上同时执行,或者这些框有时可以以相反的顺序执行,这取决于所涉及的功能。还将注意,框图和/或流程图图示的每个框以及框图和/或流程图图示中的框的组合可以由执行指定功能或动作或执行专用硬件和计算机指令的组合作用的基于硬件的系统来实现。

[0064] 图1的计算组件,包括计算机系统100,可以在一个或多个计算机系统中实现。计算机系统/服务器902可以在计算机系统可执行指令的一般上下文中描述,诸如由计算机系统执行的程序模块。通常,程序模块可以包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、逻辑、数据结构等。计算机系统/服务器902可以在分布式云计算环境中实践,其中任务由通过通信网络链接的远程处理设备执行。在分布式云计算环境中,程序模块可以位于包括存储器存储设备的本地和远程计算机系统存储介质中。

[0065] 如图9所示,计算机系统/服务器902以通用计算设备的形式示出。计算机系统/服务器902的组件可以包括但不限于一个或多个处理器或处理单元904、系统存储器906、以及将包括系统存储器906在内的各种系统组件耦合至处理器904的总线908。总线908表示若干类型的总线结构中的任何一种的一个或多个,包括存储器总线或存储器控制器、外围总线、加速图形端口、以及使用各种总线体系结构中的任何一种的处理器或局部总线。作为示例而非限制,这些体系结构包括工业标准体系结构 (ISA) 总线、微通道体系结构 (MCA) 总线、增强型ISA (EISA) 总线、视频电子技术标准协会 (VESA) 局部总线和外围部件互连 (PCI) 总线。

[0066] 计算机系统/服务器902通常包括各种计算机系统可读介质。这样的介质可以是可由计算机系统/服务器902访问的任何可用介质,并且它包括易失性和非易失性介质、可移动和不可移动介质。

[0067] 系统存储器906可以包括易失性存储器形式的计算机系统可读介质,诸如随机存取存储器 (RAM) 910和/或高速缓存912。计算机系统/服务器902还可以包括其它可移动/不可移动、易失性/非易失性计算机系统存储介质。仅作为示例,存储系统913可以被提供用于从不可移动、非易失性磁介质(未示出,并且通常被称为“硬盘驱动器”)读取和向其写入。尽管未示出,但是可以提供用于从可移动、非易失性磁盘(例如,“软盘”)读取和向其写入的磁盘驱动器,以及用于从诸如CD-ROM、DVD-ROM或其它光学介质等可移动、非易失性光盘读取或向其写入的光盘驱动器。在这样的实例中,各自可以通过一个或多个数据介质接口连接到总线908。如下面将进一步描绘和描述的,存储器906可以包括至少一个程序产品,该程序产品具有一组(例如,至少一个)程序模块,该程序模块被配置成执行本发明的实施例的功能。

[0068] 具有一组(至少一个)程序模块916的程序/实用程序914,以及操作系统、一个或多个应用程序、其它程序模块和程序数据可被存储在存储器906中,作为示例而非限制。操作系统、一个或多个应用程序、其它程序模块和程序数据或其某种组合中的每一个可包括联网环境的实现。计算机系统902的组件可以被实现为程序模块916,其通常执行如本文所描述的本发明的实施例的功能和/或方法。图1的系统可以在一个或多个计算机系统902中实现,其中如果它们在多个计算机系统902中实现,则计算机系统可以通过网络通信。

[0069] 计算机系统/服务器902还可以与一个或多个外部设备918通信,一个或多个外部设备918诸如键盘、定点设备、显示器920等,使得用户能够与计算机系统/服务器902交互的一个或多个设备,和/或使计算机系统/服务器902能够与一个或多个其它计算设备通信的任何设备(例如,网卡、调制解调器等)。这种通信可以经由输入/输出 (I/O) 接口922发生。此外,计算机系统/服务器902可以经由网络适配器924与诸如局域网 (LAN)、通用广域网 (WAN) 和/或公共网络(例如,因特网)之类的一个或多个网络通信。如所描绘的,网络适配器924经由总线908与计算机系统/服务器902的其他组件通信。应当理解,尽管未示出,但是其他硬

件和/或软件组件可以与计算机系统/服务器902结合使用。示例包括但不限于微代码、设备驱动器、冗余处理单元、外部磁盘驱动器阵列、RAID系统、磁带驱动器和数据存档存储系统等。

[0070] 术语“实施例”、“多个实施例”、“该实施例”、“这些实施例”、“一个或多个实施例”、“一些实施例”和“一个实施例”表示“本发明的一个或多个(但不是全部)实施例”,除非另有明确说明。

[0071] 除非另有明确说明,否则术语“包括”、“包含”、“具有”及其变体表示“包括但不限于”。

[0072] 除非另有明确说明,否则枚举的项的列表并不意味着任何或所有项是互斥的。

[0073] 除非另有明确说明,否则术语“一”、“一个”和“该”表示“一个或多个”。

[0074] 除非另有明确说明,否则彼此通信的设备不需要彼此连续通信。另外,彼此通信的设备可以直接通信或通过一个或多个中间物间接通信。

[0075] 具有彼此通信的若干组件的实施例的描述并不暗示需要所有此类组件。相反,描述了各种可选组件以说明本发明的各种可能的实施例。

[0076] 当本文描述单个设备或产品时,很明显,可以使用多于一个的设备/产品(无论它们是否协作)来代替单个设备/产品。类似地,在本文描述了多于一个设备或产品(无论它们是否协作)的情况下,将容易明白的是,可以使用单个设备/产品来代替多于一个设备或产品,或者可以使用不同数量的设备/产品来代替所示数量的设备或程序。设备的功能和/或特征可以替代地由一个或多个未明确描述为具有这样的功能/特征的其他设备来体现。因此,本发明的其它实施例不需要包括该设备本身。

[0077] 为了说明和描述的目的,已经给出了本发明的各种实施例的上述描述。其不是穷举的,也不是要将本发明限制为所公开的精确形式。根据上述教导,许多修改和变化是可能的。本发明的范围不是由该详细描述来限定,而是由所附权利要求来限定。上述说明、示例和数据提供了对本发明的组成的制造和使用的完整描述。由于在不脱离本发明的范围的情况下可以做出本发明的许多实施例,因此本发明在于所附权利要求书。

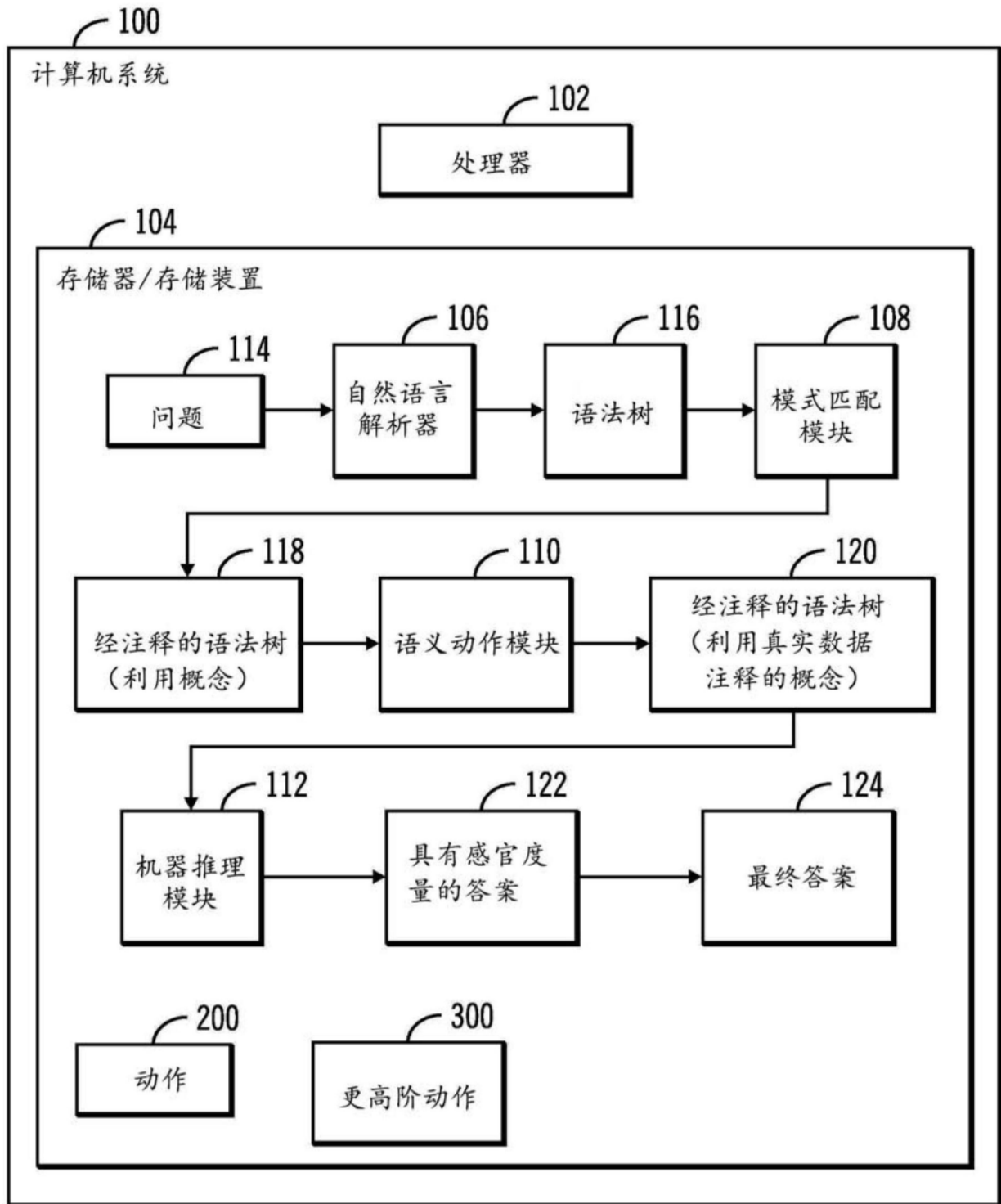


图1

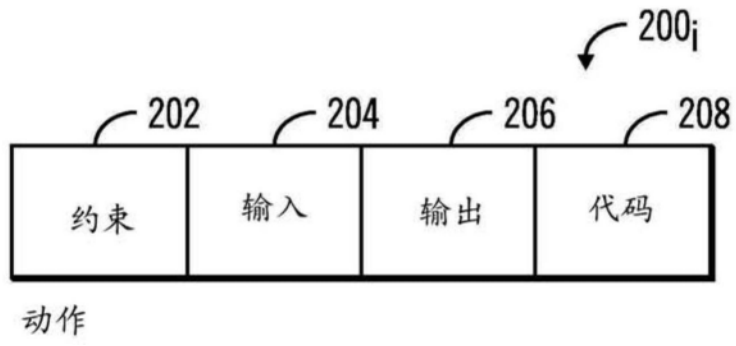


图2

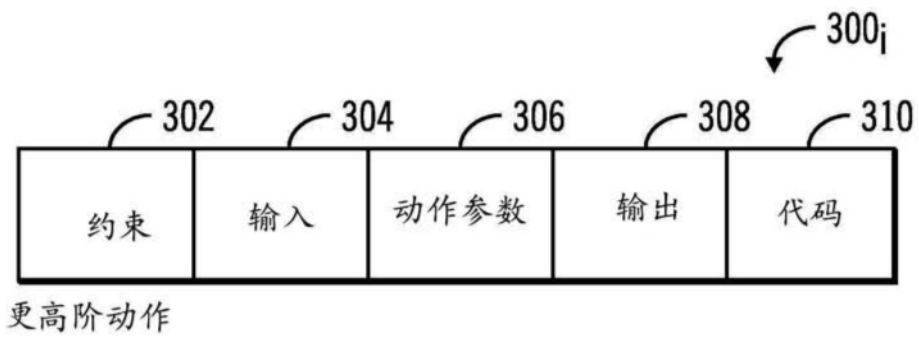


图3

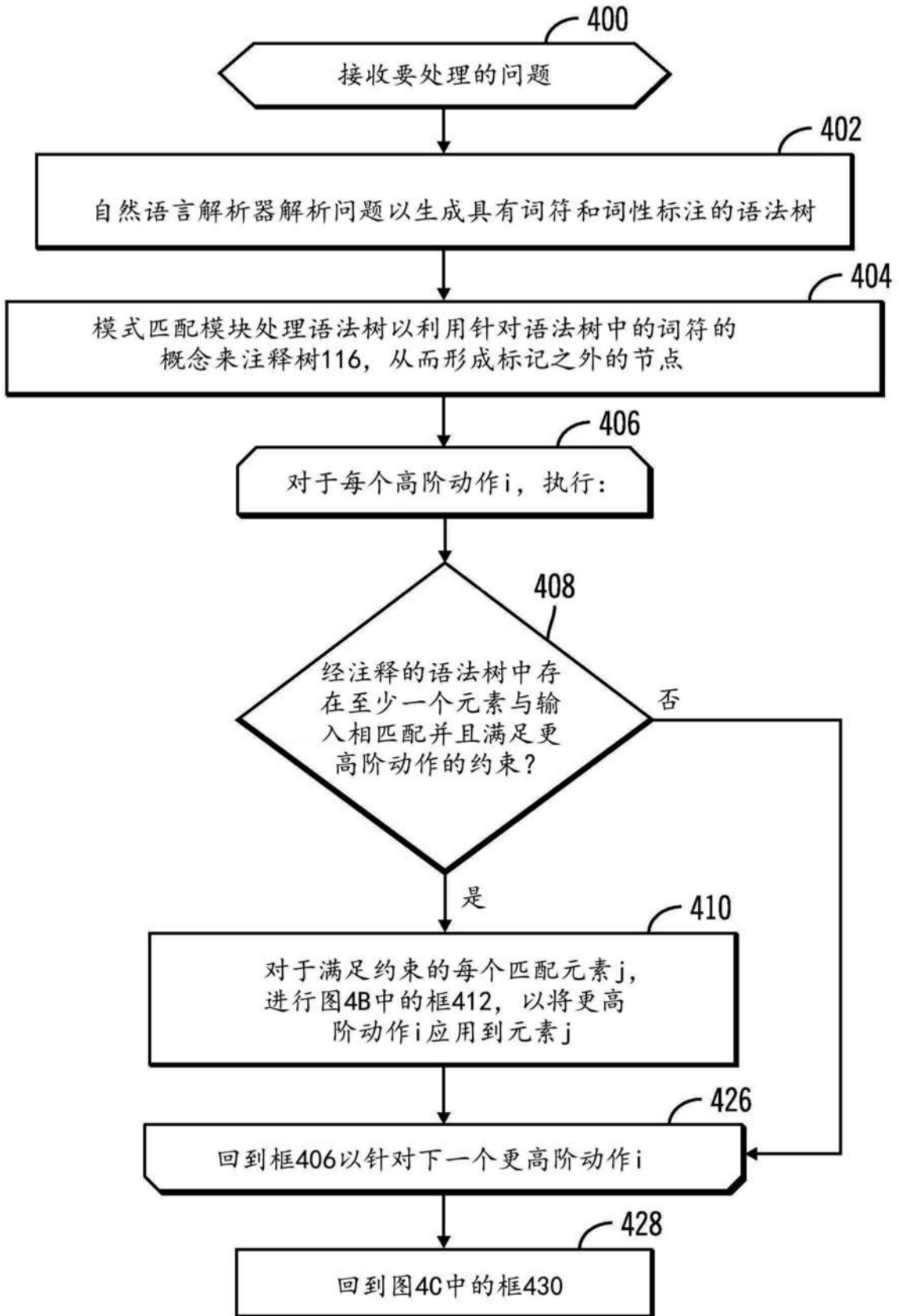


图4A

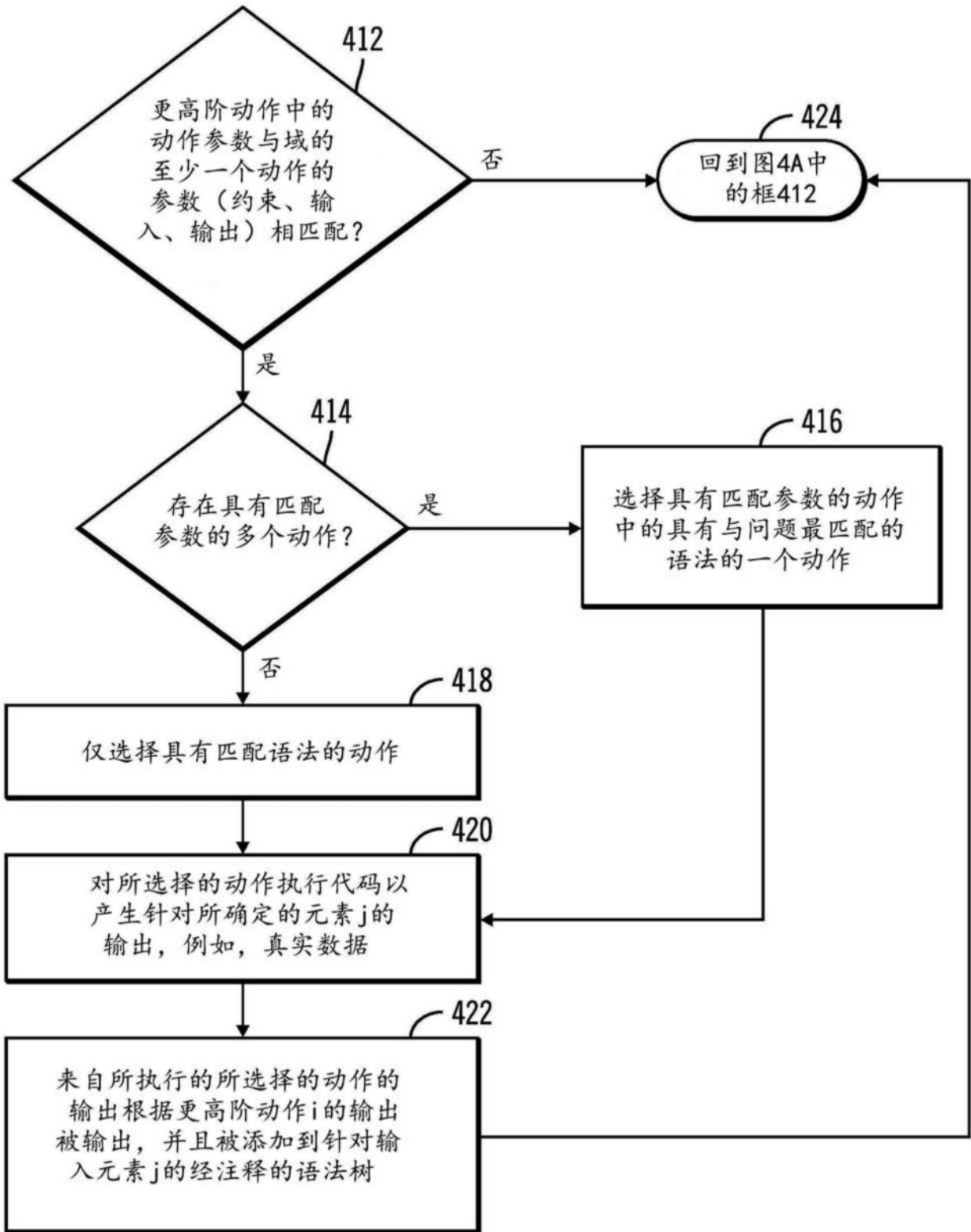


图4B

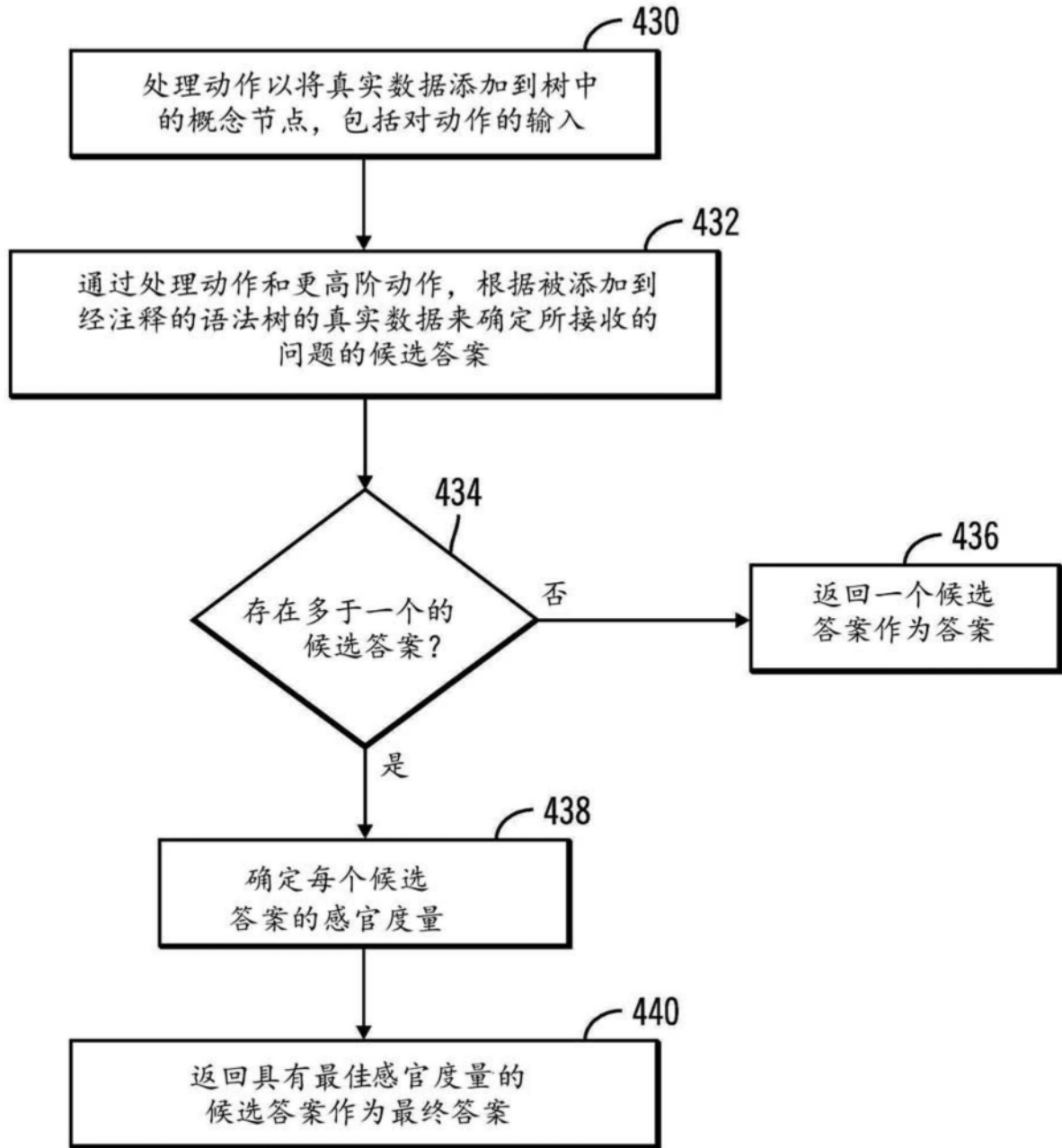


图4C

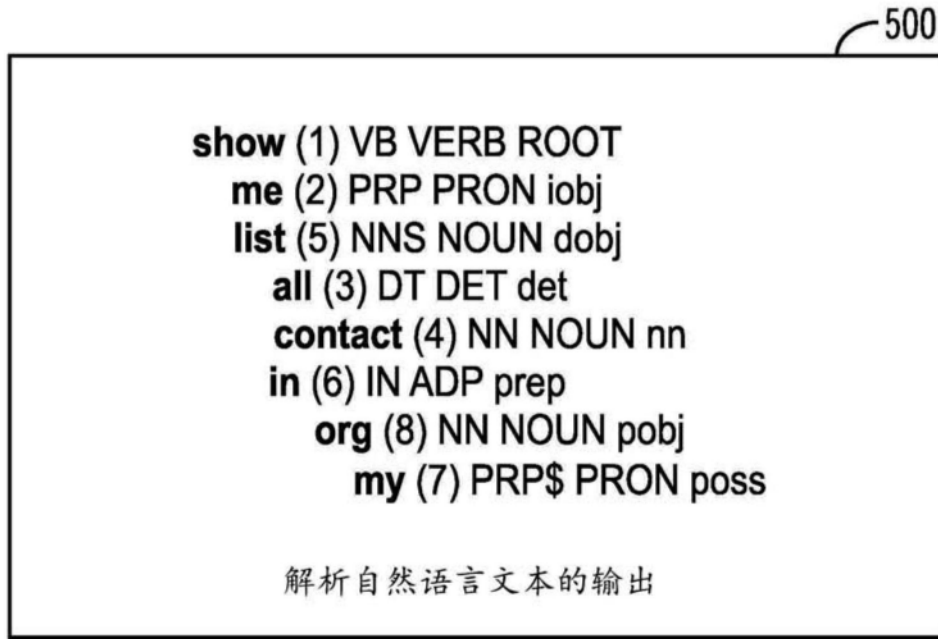


图5

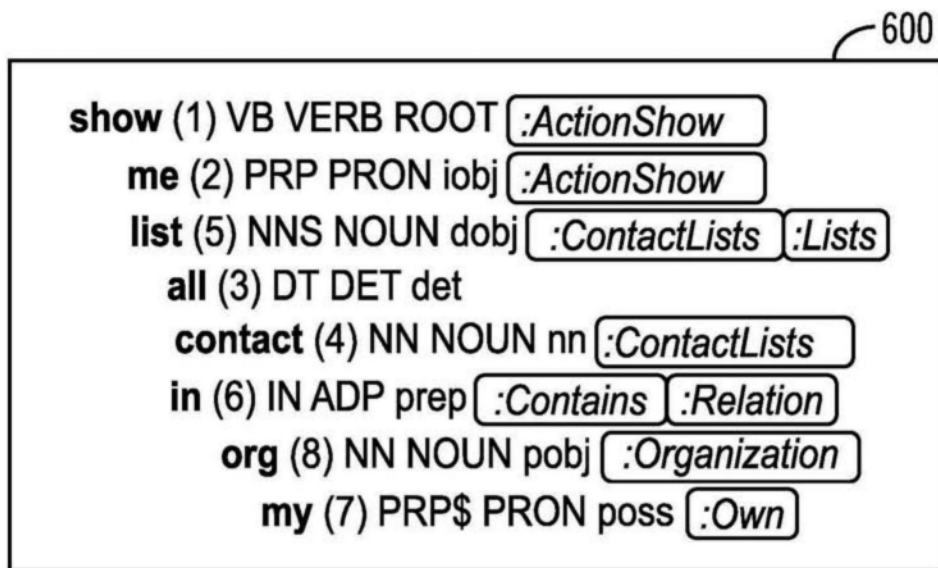


图6

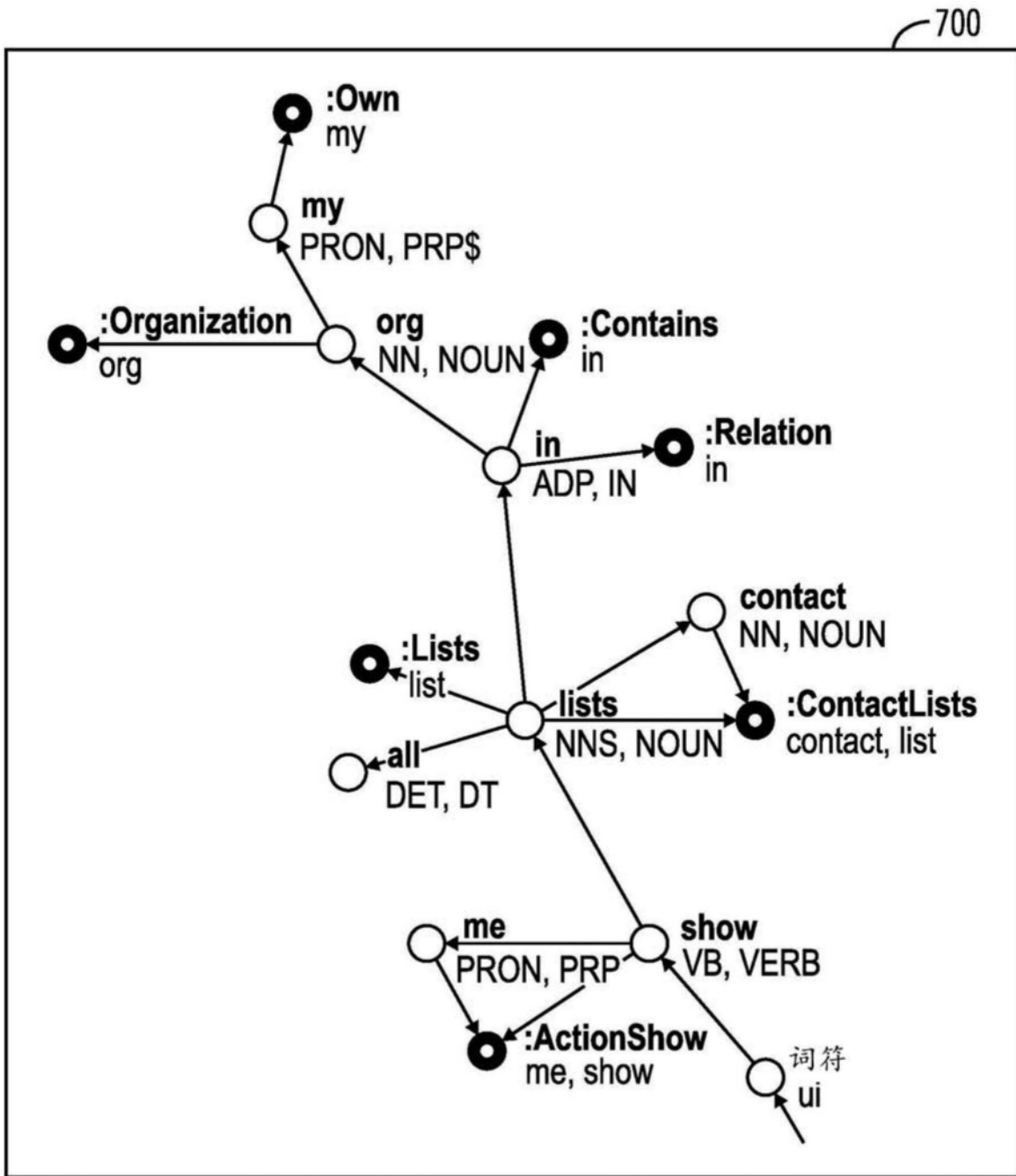


图7

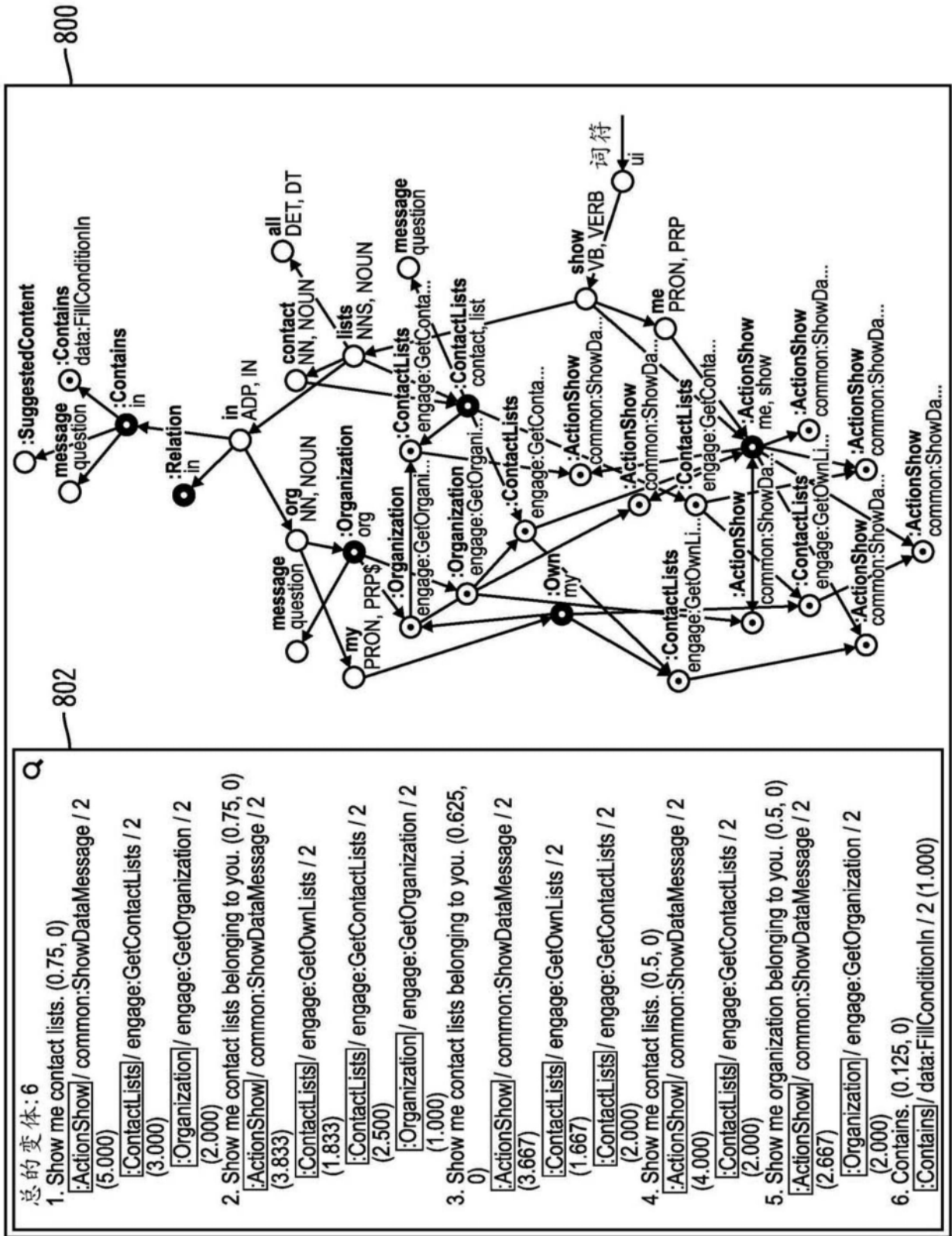


图8

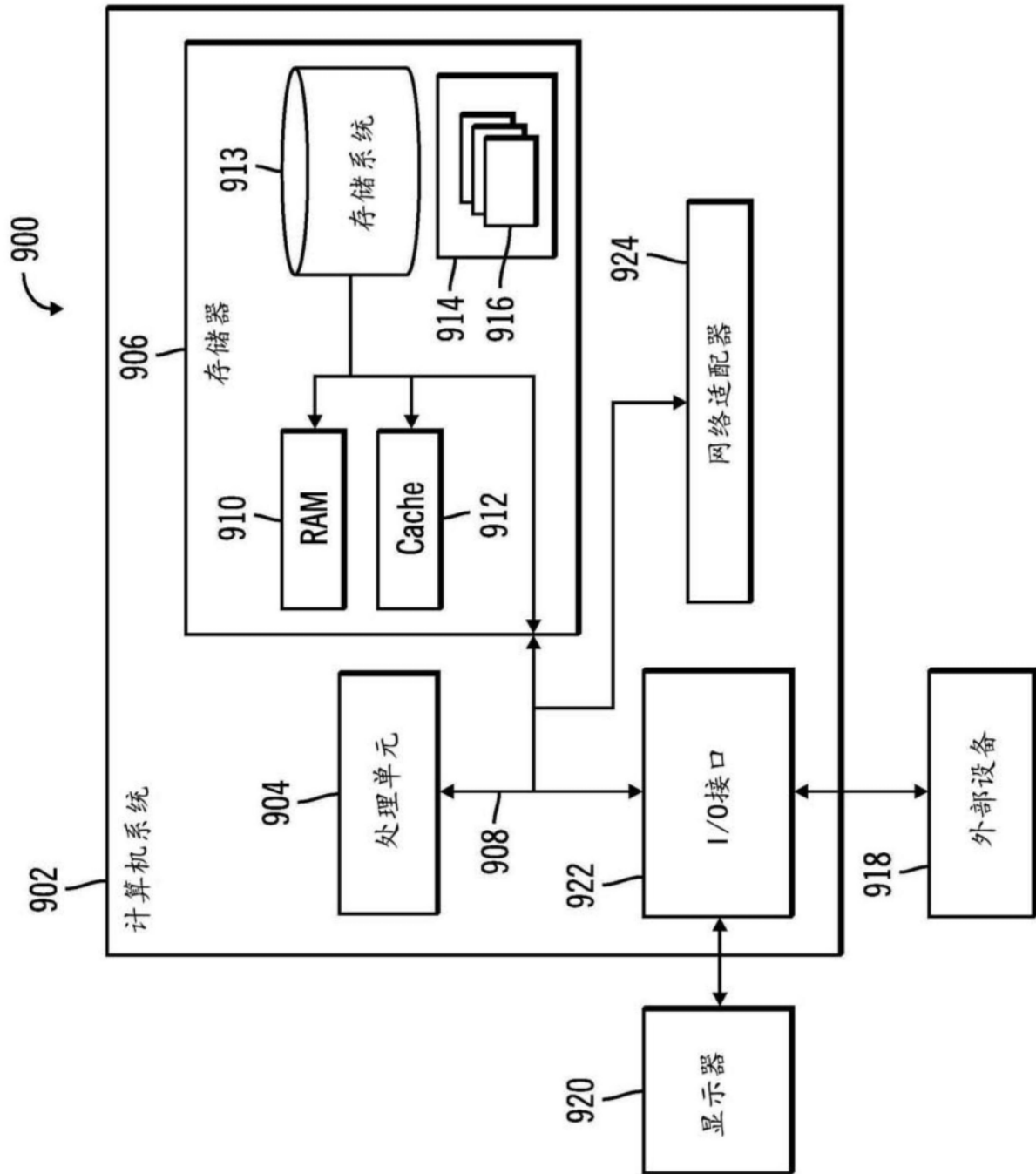


图9